

STIC-Biotech/ChemLib

79512

From: Davis, Minh-Tam
Sent: Tuesday, November 05, 2002 4:18 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/899569

SEQ ID NO: 3 and 4
Thank you.
MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

RECEIVED
NOV - 5 2002
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

PD 17-JAN-2002.
 XX
 PF 05-JUL-2001; 2001WO-EP07705.
 XX
 PR 07-JUL-2000; 2000DE-1033080.
 PR 19-APR-2001; 2001DE-1019294.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 PI Schwelger N, Scherl-Mostageer M, Sommergruber W, Abseher R:
 XX WPI: 2002-171704/22.
 DR P-PSDB; AAM49641.
 XX
 PT New tumor-associated antigen B345, useful for diagnosis and
 XX immunotherapy of tumors, also related nucleic acid and antibodies
 PS Claim 3; Page 79-85; 102pp; German.
 XX
 CC This invention describes a novel tumour-associated antigen, designated
 CC B345 which has cytostatic activity. B345 is involved in communication,
 CC interaction and/or signal transduction with extracellular components and
 CC ligands, especially in the metastatic potential of cancers, particularly
 CC of the colon. B345 or its immunogenic fragments, also the DNA that
 CC encodes it, are useful for immunotherapy of cancer, particularly
 CC treatment of lung or colon. Antibodies raised against B345 are useful for
 CC expression and diagnosis of cancers that are associated with B345
 CC radiolabeled agents. Probes derived from B345 can be used to detect
 CC tumour-specific mutations in the B345 sequence, and can be used to screen
 CC for B345 specific modulators. This sequence encodes a human B345
 CC tumour-associated antigen described in the invention.
 CC
 XX
 SQ Sequence 6163 BP; 1501 A; 1689 C; 1482 G; 1491 T; 0 other;
 Query Match 100.0%; Score 6163; DB 24; Length 6163;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ATGTGTTCATCAAGTCTGAGAAAAGATAGTCTTTACCTTTAGCTCCAGAGTCTCGAG 540
 QY AATCACTTTGTCTATAGAGATCCAGAAAAATATTGACTGTATGTCAGGCCCATGCTCTTTT 600
 DB 541 AATCACTTTGTCTATAGAGATCCAGAAAAATATTGACTGTATGTCAGGCCCATGCTCTTTT 600
 QY 601 GGGAGGTTGAGCTTCAAGCCCTGACATCGTTGTGCTTACCTCAACAGAACTTTATC 660
 DB 601 GGGAGGTTGAGCTTCAAGCCCTGACATCGTTGTGCTTACCTCAACAGAACTTTATC 660
 QY 661 TGGAGTCAAAAGTCTAAGAGAGATGTTTGTAGCTGAGTTTTCATCCCTGCTG 720
 DB 661 TGGAGTCAAAAGTCTAAGAGAGATGTTTGTAGCTGAGTTTTCATCCCTGCTG 720
 QY 721 AGGAGATCCGCTCGGCTGAGAGCTGCGCAGAGAGTCACTCACTCATCAGCGCGGA 780
 DB 721 AGGAGATCCGCTCGGCTGAGAGCTGCGCAGAGAGTCACTCACTCATCAGCGCGGA 780
 QY 781 ATGATCCCAACCGTGTGTCAGAGATCGAAGCTTCTGCAAGCAATGGCACTGTGCTCGGATC 840
 DB 781 ATGATCCCAACCGTGTGTCAGAGATCGAAGCTTCTGCAAGCAATGGCACTGTGCTCGGATC 840
 QY 841 AAGATGCAAGAGAGAGTGAATAAGGCTTACCTCCATGATGTTCCACCCAGAAATGTC 900
 DB 841 AAGATGCAAGAGAGAGTGAATAAGGCTTACCTCCATGATGTTCCACCCAGAAATGTC 900
 QY 901 TCCGCTTTCAGCATTTGCAAAACCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTG 960
 DB 901 TCCGCTTTCAGCATTTGCAAAACCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTG 960
 QY 961 TTTGAGGTTGAAGGCTTGAAGAACCTTATGTCTGCAACTTACCAAGAGCTTCCCTGAG 1020
 DB 961 TTTGAGGTTGAAGGCTTGAAGAACCTTATGTCTGCAACTTACCAAGAGCTTCCCTGAG 1020
 QY 1021 GATGAGCTCATGAGAGTGGAGTTTGTGCTCTGTCAGACCTGGGAGGAGGCTCTTTC 1080
 DB 1021 GATGAGCTCATGAGAGTGGAGTTTGTGCTCTGTCAGACCTGGGAGGAGGCTCTTTC 1080
 QY 1081 CTCAACTTCAACCTCTTCAACCTGTGAGAGAGAGAGGAGGAGGCTTGAATACTACATCCG 1140
 DB 1081 CTCAACTTCAACCTCTTCAACCTGTGAGAGAGAGAGGAGGAGGCTTGAATACTACATCCG 1140
 QY 1141 GGTCTCCACCAACCAACCCGAGGTGTTAACTGTGAGAGACAAGCAAGCTTGGGAACATGGCG 1200
 DB 1141 GGTCTCCACCAACCAACCCGAGGTGTTAACTGTGAGAGACAAGCAAGCTTGGGAACATGGCG 1200
 QY 1201 GGGAACTTCAACCTCTTCAACCTGTGAGAGAGTGTGACCAAGTCCCAAGGATCTCTC 1260
 DB 1201 GGGAACTTCAACCTCTTCAACCTGTGAGAGAGTGTGACCAAGTCCCAAGGATCTCTC 1260
 QY 1261 CGGCTGCAAGTTTCCAAAGTTTGTGTCACCAATCCAAATGAAGCAATTAATCTACGTG 1320
 DB 1261 CGGCTGCAAGTTTCCAAAGTTTGTGTCACCAATCCAAATGAAGCAATTAATCTACGTG 1320
 QY 1321 GTTGACTTGAATATGAGCAGAGCATGTCACTCACTCAAGCACCAGCCGCTCAAAACAG 1380
 DB 1321 GTTGACTTGAATATGAGCAGAGCATGTCACTCACTCAAGCACCAGCCGCTCAAAACAG 1380
 QY 1381 AGCGGCAAGTTTGTGCTGCTGTTGCTGTGATGAATCTGGAGCTGAGTAGAGCAAC 1440
 DB 1381 AGCGGCAAGTTTGTGCTGCTGTTGCTGTGATGAATCTGGAGCTGAGTAGAGCAAC 1440
 QY 1441 CTCACCTGACATCTGCTCCAAACACAAATATCTCTCTTGTGTGATGATCTGACAGT 1500
 DB 1441 CTCACCTGACATCTGCTCCAAACACAAATATCTCTCTTGTGTGATGATCTGACAGT 1500
 QY 1501 CTGTGATGATGATGAGAAAAACCATTAAGCTGACAGACACCGGTACTGCAAAAGGAAA 1560
 DB 1501 CTGTGATGATGATGAGAAAAACCATTAAGCTGACAGACACCGGTACTGCAAAAGGAAA 1560
 QY 1561 TCTTACTCATCTCAGAGTGGCCAGTGAATCTCAGTCTGCTGTGAGCTGATGATCTTC 1620
 DB 1561 TCTTACTCATCTCAGAGTGGCCAGTGAATCTCAGTCTGCTGTGAGCTGATGATCTTC 1620

Oy	1621	TCCTGGAAAGCTGCTGGTGGTCCCAAGAGACAGGCTGACCTTGCTGTGTGTCGACGCCAGAG	1680
Dp	1621	TCTCGGAAGGCTGCTGGTGGTCCCAAGAGACAGGCTTCAGCTGTGTGTGTCGACGCCAGAG	1680
Oy	1661	CTGAGAGAGCATACACAGAGAAAGCCCTGCAACACAGCTTAGAGTACTGTGTGGCCAGT	1740
Dp	1661	CTGAGAGAGCATACACAGAGAAAGCCCTGCAACACAGCTTAGAGTACTGTGTGGCCAGT	1740
Oy	1741	GCCATTACCCAGCCAGGACCTGTAACCTTCGAGCTCCTTGTGCCGGAGAGGCTATCAAGAG	1800
Dp	1741	GCCATTACCCAGCCAGGACCTGTAACCTTCGAGCTCCTTGTGCCGGAGAGGCTATCAAGAG	1800
Oy	1801	ATCCAGGTGAAGCAGAAACATCTCGGTGACCCCTTCGACCTTTGCCCCAGCTTCCACAA	1860
Dp	1801	ATCCAGGTGAAGCAGAAACATCTCGGTGACCCCTTCGACCTTTGCCCCAGCTTCCACAA	1860
Oy	1861	GAGGCTTCACAGGAGGCTGTGAGGGTGTCTTTATACCTATTATTCAAAGAGAGGCTT	1920
Dp	1861	GAGGCTTCACAGGAGGCTGTGAGGGTGTCTTTATACCTATTATTCAAAGAGAGGCTT	1920
Oy	1921	TTACAGGAGACCCCTGACACAAAAAGAGAGTACTACCTAGAGACCCCAACTGGGAGACGG	1980
Dp	1921	TTACAGGAGACCCCTGACACAAAAAGAGAGTACTACCTAGAGACCCCAACTGGGAGACGG	1980
Oy	1981	GGCCTGCCATCCCTCACCTGTGTCTCTGGAACATCAGCGTGCACAGAGACAGGTGGCC	2040
Dp	1981	GGCCTGCCATCCCTCACCTGTGTCTCTGGAACATCAGCGTGCACAGAGACAGGTGGCC	2040
Oy	2041	TGCCGTACTTCTTTAAAGAGCGGAGCGGCTGTGTCACAGACAGGCGCGCATTCATG	2100
Dp	2041	TGCCGTACTTCTTTAAAGAGCGGAGCGGCTGTGTCACAGACAGGCGCGCATTCATG	2100
Oy	2101	ATCATCCAGAGGACGCGGACCCGGGCGAGAGATCTTAGCGCTGACAGAGATGTGTC	2160
Dp	2101	ATCATCCAGAGGACGCGGACCCGGGCGAGAGATCTTAGCGCTGACAGAGATGTGTC	2160
Oy	2161	CCCAAGCCAAACCTTCCACCATCAGAGCTTCTGGGTACACATCTTAACGTGACGCCACAG	2220
Dp	2161	CCCAAGCCAAACCTTCCACCATCAGAGCTTCTGGGTACACATCTTAACGTGACGCCACAG	2220
Oy	2221	AGCGGCAAGACGCTAGACCTGTCTTCTCGGTGACACTTAACCCAAAGGACTGTGACCTTG	2280
Dp	2221	AGCGGCAAGACGCTAGACCTGTCTTCTCGGTGACACTTAACCCAAAGGACTGTGACCTTG	2280
Oy	2281	ACTGTCATCCATCGGAGGCGGAGGGTGGAGGCTTACTGCTGTGTCCCTCGGGCTC	2340
Dp	2281	ACTGTCATCCATCGGAGGCGGAGGGTGGAGGCTTACTGCTGTGTCCCTCGGGCTC	2340
Oy	2341	ATCATTTGCTGTGTGAAGAAAGAAAGAAAGAACAAACAAAGGCGCTGTGGGTATC	2400
Dp	2341	ATCATTTGCTGTGTGAAGAAAGAAAGAAAGAACAAACAAAGGCGCTGTGGGTATC	2400
Oy	2401	TACAAATGGCAACATCAATACTGAGATGCGGAGCGACCCAAAAAAGTTTCAGAAAGGCGGA	2460
Dp	2401	TACAAATGGCAACATCAATACTGAGATGCGGAGCGACCCAAAAAAGTTTCAGAAAGGCGGA	2460
Oy	2461	AAGGACAATGATCCCATGTGTATGCAAGTCATGAGAGAACCATGATATATGGGATCTG	2520
Dp	2461	AAGGACAATGATCCCATGTGTATGCAAGTCATGAGAGAACCATGATATATGGGATCTG	2520
Oy	2521	CTACAGAGATTCCAGCGGCTCTTCTCTGACAGCCAGAGGTGGACACTACCGGCGCTTCCAG	2580
Dp	2521	CTACAGAGATTCCAGCGGCTCTTCTCTGACAGCCAGAGGTGGACACTACCGGCGCTTCCAG	2580
Oy	2581	GGCAACCATGGGGGTCTGTCCCTCCCTCCCAACCCACCATATGCTCCAGGCGCCCAACTGCA	2640
Dp	2581	GGCAACCATGGGGGTCTGTCCCTCCCTCCCAACCCACCATATGCTCCAGGCGCCCAACTGCA	2640
Oy	2641	AAGTTGGGCACCTGAGAGAGCAACTCTCGCTCCCTCTCTGAGTCTGAGAGTGAACCGTAC	2700
Dp	2641	AAGTTGGGCACCTGAGAGAGCAACTCTCGCTCCCTCTCTGAGTCTGAGAGTGAACCGTAC	2700

QY	2701	ACCTCTCCCAATCCCAACAAATGGGAGATGTAAAGCAGCAAGAGACAGACATTCCTCTACTG	2760
Db	2701	ACCTCTCTCCCAATCCCAACAAATGGGAGATGTAAAGCAGCAAGAGACAGACATTCCTCTACTG	2760
QY	2761	AACACTCAGGAGCCCATGAGGCCAGGACGAAATTAATTGATATCCAAATTCAGAGCTTTGGCTGA	2820
Db	2761	AACACTCAGGAGCCCATGAGGCCAGGACGAAATTAATTGATATCCAAATTCAGAGCTTTGGCTGA	2820
QY	2821	GTTCATTAAGCAGGAGCACTGAGACACCCGCTCCGTTCTCTTAACAGAAATCTTAAGAA	2880
Db	2821	GTTCATTAAGCAGGAGCACTGAGACACCCGCTCCGTTCTCTTAACAGAAATCTTAAGAA	2880
QY	2881	GAGGAATTAATACAGAAAGAACAGCAGAGGTTTTCTGACACACGCCAACTTCACATTCG	2940
Db	2881	GAGGAATTAATACAGAAAGAACAGCAGAGGTTTTCTGACACACGCCAACTTCACATTCG	2940
QY	2941	TCAGGAGCACTAATCTTAAGGGCAAGACATGAAATGAATGAATTCCAATCTGGATACAGT	3000
Db	2941	TCAGGAGCACTAATCTTAAGGGCAAGACATGAAATGAATGAATTCCAATCTGGATACAGT	3000
QY	3001	CATGACACCTCATGTGCTGCCCAACTAGGAGCTGTGGGTAGCCAGCCGTAATGAGAG	3060
Db	3001	CATGACACCTCATGTGCTGCCCAACTAGGAGCTGTGGGTAGCCAGCCGTAATGAGAG	3060
QY	3061	AGAGAGGCCCTGACCTACCTAGCATAGGGTTGACAGAACCCCTGAGATTCCAGAGTGTAAAC	3120
Db	3061	AGAGAGGCCCTGACCTACCTAGCATAGGGTTGACAGAACCCCTGAGATTCCAGAGTGTAAAC	3120
QY	3121	AGAGGCTTGCCCTCTTACAGGACAAAGTTCCAAATTCCAAGAGCCTACTGAGGTCCTTA	3180
Db	3121	AGAGGCTTGCCCTCTTACAGGACAAAGTTCCAAATTCCAAGAGCCTACTGAGGTCCTTA	3180
QY	3181	CTCTACTGGGGGTCCCAAGATGAAACGACAAATGTGCCCTTTTATATTAATTAATTTTGG	3240
Db	3181	CTCTACTGGGGGTCCCAAGATGAAACGACAAATGTGCCCTTTTATATTAATTAATTTTGG	3240
QY	3241	TGGTCTGTGTTATTAAGAGATCAAAATGTATAACACACTAGCTCTTTACCTACTACTTA	3300
Db	3241	TGGTCTGTGTTATTAAGAGATCAAAATGTATAACACACTAGCTCTTTACCTACTACTTA	3300
QY	3301	GTAATACTCTACTAATCTGGTTGGATGGCTGGGTTGGAGCTTCTACAGACCCGTAAT	3360
Db	3301	GTAATACTCTACTAATCTGGTTGGATGGCTGGGTTGGAGCTTCTACAGACCCGTAAT	3360
QY	3361	AAACGTGCGCTGTCCCAAGAGTGGTGGAAATTAATTAACAACTGTCCAAACAGAAAAGA	3420
Db	3361	AAACGTGCGCTGTCCCAAGAGTGGTGGAAATTAATTAACAACTGTCCAAACAGAAAAGA	3420
QY	3421	ATGTGTGTTGTGAGCAGACATTCACACATATGCTTTGATAAAGACTTCTGTATTCYC	3480
Db	3421	ATGTGTGTTGTGAGCAGACATTCACACATATGCTTTGATAAAGACTTCTGTATTCYC	3480
QY	3481	TAGGTCGGTTCGTGTTATATCCATATGTGGAAATTCATCTTGAATCCCATTTGCTCATAGT	3540
Db	3481	TAGGTCGGTTCGTGTTATATCCATATGTGGAAATTCATCTTGAATCCCATTTGCTCATAGT	3540
QY	3541	CCTAGCAATTAAGAAATTTCCTAATTTCCATGTGCGGTTCTCCTAGCTCAGCAATA	3600
Db	3541	CCTAGCAATTAAGAAATTTCCTAATTTCCATGTGCGGTTCTCCTAGCTCAGCAATA	3600
QY	3601	CTTTGACATTTAAGAGAAATTTAGAAATATTCATCCTCTAAAAATGTTTAATATA	3660
Db	3601	CTTTGACATTTAAGAGAAATTTAGAAATATTCATCCTCTAAAAATGTTTAATATA	3660
QY	3661	TACCAAAAGTGGCCCTGCTAGTTAGTTTCTGTGTGACATGCAACCAATTACTTGGTAG	3720
Db	3661	TACCAAAAGTGGCCCTGCTAGTTAGTTTCTGTGTGACATGCAACCAATTACTTGGTAG	3720
QY	3721	CTTAATAAACACACATTAAGCTTAATAGTCGTGGGGAATCAGAAATTCCAAATGGATGCCCT	3780
Db	3721	CTTAATAAACACACATTAAGCTTAATAGTCGTGGGGAATCAGAAATTCCAAATGGATGCCCT	3780
QY	3781	GAATGAAAAATCAGGTGTACAGCAGAGTGTGCTCTTCTGAAGAGCTCTAGGAGAAAGCG	3840

Db 3781 GAAAGAAATCAAGGTGACAGAGGCTGCTCTTCTGAAAGCTCTAGGAAAGCG 3840
Qy 3841 GTTCCTTGCCATTCAAGCTCTAGAGGCTGGCTGCAATTCACAGGCTCCAGTGGCTG 3900
Db 3841 GTTCCTTGCCATTCAAGCTCTAGAGGCTGGCTGCAATTCACAGGCTCCAGTGGCTG 3900
Qy 3901 AAGCTTTCTCAGATGCGATCTGTGACACATGAGCCCTCCACATCTCTTTGACTTAC 3960
Db 3901 AAGCTTTCTCAGATGCGATCTGTGACACATGAGCCCTCCACATCTCTTTGACTTAC 3960
Qy 3961 AAAGCCACCGAAGATCCAGGATCTCTCATCTAAAGATCTTCAATCCCTGGA 4020
Db 3961 AAAGCCACCGAAGATCCAGGATCTCTCATCTAAAGATCTTCAATCCCTGGA 4020
Qy 4021 AGAGCTTTTGCCATGCAAGACATAGCACAGTGGGGATTAGGACAGAGATCTT 4080
Db 4021 AGAGCTTTTGCCATGCAAGACATAGCACAGTGGGGATTAGGACAGAGATCTT 4080
Qy 4081 TGGGGTCTGTTATTTGCTTACCAACATCTTCCGCACTGATCCACAGAGAGCTA 4140
Db 4081 TGGGGTCTGTTATTTGCTTACCAACATCTTCCGCACTGATCCACAGAGAGCTA 4140
Qy 4141 CAAATGATCGGGGACAGGAGATGTTTGTAGCTTGGGAGCTTAACCTTAAAAA 4200
Db 4141 CAAATGATCGGGGACAGGAGATGTTTGTAGCTTGGGAGCTTAACCTTAAAAA 4200
Qy 4201 ACCCCAGATCAGAAAGATCTGGCCATGCTGGGGCTCACATTTCCACTTACAGCAACTGGC 4260
Db 4201 ACCCCAGATCAGAAAGATCTGGCCATGCTGGGGCTCACATTTCCACTTACAGCAACTGGC 4260
Qy 4261 TGGAGCTGGGACACAGCTCTGCTTTAGAAAGGGGTCTCAGCTTACAGAGTCCACAGC 4320
Db 4261 TGGAGCTGGGACACAGCTCTGCTTTAGAAAGGGGTCTCAGCTTACAGAGTCCACAGC 4320
Qy 4321 CCAACTACGCGCTATCACTCCACATAGAGGCTAGTGTCTTCTACTGATCATG 4380
Db 4321 CCAACTACGCGCTATCACTCCACATAGAGGCTAGTGTCTTCTACTGATCATG 4380
Qy 4381 CCCCTGCAAGTGTGATTTATGAAAAAAGAAAGTGGGATATCTCTATCAGG 4440
Db 4381 CCCCTGCAAGTGTGATTTATGAAAAAAGAAAGTGGGATATCTCTATCAGG 4440
Qy 4441 TGAATACCATGAGACCAATGTGCTCACATTAACCTTTCTTTTCTTTTCT 4500
Db 4441 TGAATACCATGAGACCAATGTGCTCACATTAACCTTTCTTTTCTTTTCT 4500
Qy 4501 TTTTCTTTTCTTTTATGAGAGAGATCTCATCTGTTGCTAGGCTGAGGCTCAG 4560
Db 4501 TTTTCTTTTCTTTTATGAGAGAGATCTCATCTGTTGCTAGGCTGAGGCTCAG 4560
Qy 4561 TGGGCAATCTCGGCTCACTGCAACCTCTGCTGCTGGGCTCAAGCAATTTCCACCTC 4620
Db 4561 TGGGCAATCTCGGCTCACTGCAACCTCTGCTGCTGGGCTCAAGCAATTTCCACCTC 4620
Qy 4621 AGCTTCCCAATAGCTGGATCACTGCGACAACACCATGCCCCAGCTAATTTGTATT 4680
Db 4621 AGCTTCCCAATAGCTGGATCACTGCGACAACACCATGCCCCAGCTAATTTGTATT 4680
Qy 4681 TTTGTAGAGAGGCTTCAACATGTTGCCAGGCTGCTCAACCTCTGAGGCTCAGC 4740
Db 4681 TTTGTAGAGAGGCTTCAACATGTTGCCAGGCTGCTCAACCTCTGAGGCTCAGC 4740
Qy 4741 AATCTCTGCTGCGCTCCCAAAAGTGTGAGATACAGATGAGCCACCGCATCAG 4800
Db 4741 AATCTCTGCTGCGCTCCCAAAAGTGTGAGATACAGATGAGCCACCGCATCAG 4800
Qy 4801 CCCACACCTCATTTATACCAATTTACCTGCGCAGTAAGTGTGAGCTTTGCTCTCAC 4860
Db 4801 CCCACACCTCATTTATACCAATTTACCTGCGCAGTAAGTGTGAGCTTTGCTCTCAC 4860
Qy 4861 CCTGCTCATCTGGAAGAGAGGATTTATGCTTGTGAGCAAGTCCCAAGTT 4920
Db 4861 CCTGCTCATCTGGAAGAGAGGATTTATGCTTGTGAGCAAGTCCCAAGTT 4920

Db 4861 CCTGCTCATCTGGAAGAGAGGATTTATGCTTGTGAGCAAGTCCCAAGTT 4920
Qy 4921 CAATATTTCTGCGGCAAAACTTCTTCAAAAATAATGACTTCAATTTCAATGA 4980
Db 4921 CAATATTTCTGCGGCAAAACTTCTTCAAAAATAATGACTTCAATTTCAATGA 4980
Qy 4981 ATTACCTTGAATGACACCGCTCAACTTGTTCATGATGCAATGAAGAAATTTTA 5040
Db 4981 ATTACCTTGAATGACACCGCTCAACTTGTTCATGATGCAATGAAGAAATTTTA 5040
Qy 5041 TAGTCTCTAAATGAGGCTGATCTGCAAGACCTCTGGAACCTTTCCAGAGATGATAT 5100
Db 5041 TAGTCTCTAAATGAGGCTGATCTGCAAGACCTCTGGAACCTTTCCAGAGATGATAT 5100
Qy 5101 TTAAGTATGCTTGGCGTGGCTATGAGCACTTCTCTTCAAAAGTCTGTTCTGTC 5160
Db 5101 TTAAGTATGCTTGGCGTGGCTATGAGCACTTCTCTCTTCAAAAGTCTGTTCTGTC 5160
Qy 5161 CCACTGACCTTGGCTTGTGAGCGAGATGCTGACCTTGATTAAGGGCCAAAGAGG 5220
Db 5161 CCACTGACCTTGGCTTGTGAGCGAGATGCTGACCTTGATTAAGGGCCAAAGAGG 5220
Qy 5221 CTGGGCTCTCTTCCCTCAGTGAAGAGCCCTATTGAATTCATGTTGAGGCTCAGC 5280
Db 5221 CTGGGCTCTCTTCCCTCAGTGAAGAGCCCTATTGAATTCATGTTGAGGCTCAGC 5280
Qy 5281 CTTCCATTTCTGCAATTTCCCAACCTCCAGCCCTTCCAAAGAGAGATAGTCCCTGTC 5340
Db 5281 CTTCCATTTCTGCAATTTCCCAACCTCCAGCCCTTCCAAAGAGAGATAGTCCCTGTC 5340
Qy 5341 ATTCCACCCAAAGTGGATTTGCTTCTTAAAGCCAGGCCCCCATTTACATTAACAT 5400
Db 5341 ATTCCACCCAAAGTGGATTTGCTTCTTAAAGCCAGGCCCCCATTTACATTAACAT 5400
Qy 5401 CACTGTTCCTGCAAGAGACACAGTGGCCATTTCTTCACTGAGGCTCAAACTCC 5460
Db 5401 CACTGTTCCTGCAAGAGACACAGTGGCCATTTCTTCACTGAGGCTCAAACTCC 5460
Qy 5461 TGGCAAGTGTGCTGCTCTGAGACAGATTTTCTGAGGCTGCTCTCACTGAGGAGG 5520
Db 5461 TGGCAAGTGTGCTGCTCTGAGACAGATTTTCTGAGGCTGCTCTCACTGAGGAGG 5520
Qy 5521 CCAAGCTGAGAACCTTGGCTCTTTCTTTAAAGCCAGGCCCCCATTTACATTAACAT 5580
Db 5521 CCAAGCTGAGAACCTTGGCTCTTTCTTTAAAGCCAGGCCCCCATTTACATTAACAT 5580
Qy 5581 TCAAGGTCACCTGGAACAGGAAGTGCATTTGTTGAAGCTTATGATGACAGCCACT 5640
Db 5581 TCAAGGTCACCTGGAACAGGAAGTGCATTTGTTGAAGCTTATGATGACAGCCACT 5640
Qy 5641 GCTCATCCAGTGGTCTGCCATGCTCTGAGAGAGGCGAGCATGCAAGACTGCTCT 5700
Db 5641 GCTCATCCAGTGGTCTGCCATGCTCTGAGAGAGGCGAGCATGCAAGACTGCTCT 5700
Qy 5701 AATGCTGTGCTATGCAAGAGAGGAAAGTCTCAAGAAAGTCAACTGAGCAAGCA 5760
Db 5701 AATGCTGTGCTATGCAAGAGAGGAAAGTCTCAAGAAAGTCAACTGAGCAAGCA 5760
Qy 5761 CAAGCCACCGGACATGAGGCTTGTAAAGTTAGCAACAGTGTGATGATGAGT 5820
Db 5761 CAAGCCACCGGACATGAGGCTTGTAAAGTTAGCAACAGTGTGATGATGAGT 5820
Qy 5821 GCTTCACTGGAATTAATTTATCTTCAATGAGTGGTCAATTTATTCATTTT 5880
Db 5821 GCTTCACTGGAATTAATTTATCTTCAATGAGTGGTCAATTTATTCATTTT 5880
Qy 5881 CTGCTCTTAAATTAACAAATGTAACAAAGATATCAAGCTGTTAAAGTCTTCGG 5940
Db 5881 CTGCTCTTAAATTAACAAATGTAACAAAGATATCAAGCTGTTAAAGTCTTCGG 5940
Qy 5941 CTACTGTCCCGGCTTCAATAGAGGCGCGGTTTCCCATTTGTTGACCTGACAGGCTC 6000
Db 5941 CTACTGTCCCGGCTTCAATAGAGGCGCGGTTTCCCATTTGTTGACCTGACAGGCTC 6000

07	6001	AGCATGGCGCTCAGCAGATCTCTGCTTAATTTTGCGATGATACAGAAAGCAGGCTTTGGG	6060
07	6001		
Db	6001	AGCATGGCGCTCAGCAGATCTCTGCTTAATTTTGCGATGATACAGAAAGCAGGCTTTGGG	6060
07	6061	ATACAAAGTTCTTCCCTTCATTTGATGGCGGTGCTGCTGATGAGCAGATGTTTGTCC	6120
Db	6061	ATACAAAGTTCTTCCCTTCATTTGATGGCGGTGCTGCTGATGAGCAGATGTTTGTCC	6120
07	6121	GGAATTAATAATTAATGTTGGAGTCTGCCCAAAAAAAAAA	6163
Db	6121	GGAATTAATAATTAATGTTGGAGTCTGCCCAAAAAAAAAA	6163
RESULT 2			
ABAB99506			
ID	ABA99506	standard; cDNA; 5897 BP.	
XX	ABA99506;		
XX	17-MAY-2002	(first entry)	
XX	Human tumour-associated antigen B345 cDNA.		
XX	Tumour-associated antigen; human; B345; cytostatic; cell communication;		
XX	cell interaction; signal transduction; metastasis; cancer; colon;		
XX	immunotherapy; carcinoma; lung; diagnosis; gene; ss.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
XX	5'UTR	1..214	
XX	CDS	/*tag= a	
XX		215..2464	
XX		/tag= b	
XX		/product= "tumour-associated antigen B345"	
XX	3'UTR	2465..5897	
XX		/*tag= c	
XX	WO200204508-A1.		
XX	17-JAN-2002.		
XX	.05-JUL-2001; 2001WO-EP07705.		
XX	07-JUL-2000; 2000DE-1033080.		
XX	PR 19-APR-2001; 2001DE-1019294.		
XX	(BOEH) BOEHRINGER INGELHEIM INT GMBH.		
XX	Schweifer N, Scherl-Mostlagger M, Sommergruber W, Abseher R;		
XX	WPI: 2002-171704/22.		
XX	P-PSDB; AAM49640.		
XX	New tumour-associated antigen B345, useful for diagnosis and		
XX	immunotherapy of tumors, also related nucleic acid and antibodies		
XX	Example 6; Page 70-76; 102pp; German.		
XX	This invention describes a novel tumour-associated antigen, designated		
XX	B345 which has cytostatic activity. B345 is involved in communication and		
XX	interaction and/or signal transduction with extracellular components and		
XX	ligands, especially in the metastatic potential of cancers, particularly		
XX	of the colon. B345 or its immunogenic fragments, also the DNA that		
XX	encodes it, are useful for immunotherapy of cancer, particularly		
XX	carcinoma of lung or colon. Antibodies raised against B345 are useful for		
XX	treatment and diagnosis of cancers that are associated with B345		
XX	expression, including their use for targeted delivery of cytotoxic or		
XX	radioactive agents. Probes derived from B345 can be used to detect		
XX	tumour-specific mutations in the B345 sequence, and can be used to screen		
XX	for B345 specific modulators. This sequence encodes a human B345		
XX	tumour-associated antigen described in the invention.		

Db 985 AGGATCTCTCGGGGAGTTCAGATTGTTGTCACATCTCCACAAATGAAACCAATA 1044
QY 1311 AATTCAGTGTGTTACTGTTAGTAATGAGCGACCATGTCTACTCACCATCGACCGCC 1370
Db 1045 AATTCAGTGTGTTACTGTTAGTAATGAGCGACCATGTCTACTCACCATCGACCGCC 1144
QY 1371 CGTCAACAGAGCGGCAAGTTGTCCCGGCTGTTGTGTGTAGAACTCCGACTG 1430
Db 1105 CGTCAACAGAGCGGCAAGTTGTCCCGGCTGTTGTGTGTAGAACTCCGACTG 1164
QY 1431 CAGTAGCAACCTCACCCTGACATCTGGCTCCAAACACAAATCTCTTCTTGTGATGA 1490
Db 1165 CAGTAGCAACCTCACCCTGACATCTGGCTCCAAACACAAATCTCTTCTTGTGATGA 1224
QY 1491 TCTGACACGTGTGTGATGAATGTGAAAAACATAAGCTGCACAGACACCGGTACTG 1550
Db 1225 TCTGACACGTGTGTGATGAATGTGAAAAACATAAGCTGCACAGACACCGGTACTG 1284
QY 1551 CCAAGGAAATCTTACTCCTCCAGGTGCCAGTAGACATCTCCACCTGCTGTGAGACT 1610
Db 1285 CCAAGGAAATCTTACTCCTCCAGGTGCCAGTAGACATCTCCACCTGCTGTGAGACT 1344
QY 1611 GCATGACTTCTCTGGAAGCTGTGTGTGCCAAGACAGGCTCAGCCTGGTGTGTGCC 1670
Db 1345 GCATGACTTCTCTGGAAGCTGTGTGTGCCAAGACAGGCTCAGCCTGGTGTGTGCC 1404
QY 1671 AGCCCAAGCTGCAGACGATACACAGAGAACCTCTGCAACACAGCTTCAGCTACT 1730
Db 1405 AGCCCAAGCTGCAGACGATACACAGAGAACCTCTGCAACACAGCTTCAGCTACT 1464
QY 1731 CGTGGGAGTGGCCATACCCAGGCGAGCCGTACTTGGGCTCTTCTGCCCCGGAGGCTC 1790
Db 1465 CGTGGGAGTGGCCATACCCAGGCGAGCCGTACTTGGGCTCTTCTGCCCCGGAGGCTC 1534
QY 1791 TATCAACAGATCCAGGTGGAAGCAGAACATCTCGGTGACCTTGGCACTTTGGCCCCAG 1850
Db 1325 TATCAACAGATCCAGGTGGAAGCAGAACATCTCGGTGACCTTGGCACTTTGGCCCCAG 1584
QY 1851 CTTTCAACAGAGGCTCTCCAGGAGGCTGTGACGGTCTCTTATACCTTATTTCAAGA 1910
Db 1585 CTTTCAACAGAGGCTCTCCAGGAGGCTGTGACGGTCTCTTATACCTTATTTCAAGA 1644
QY 1911 GGAAGGGGTTTTCAGGGGAGCCCTGACACAAAAAGCAAGGTCTACCTGAGGAGCCCA 1970
Db 1645 GGAAGGGGTTTTCAGGGGAGCCCTGACACAAAAAGCAAGGTCTACCTGAGGAGCCCA 1704
QY 1971 CTGGGAGCCGGGCTGACATCTCTACCTGTGTCTCTGGAACATCAGCGTGCCAGAGA 2030
Db 1705 CTGGGAGCCGGGCTGACATCTCTACCTGTGTCTCTGGAACATCAGCGTGCCAGAGA 1764
QY 2031 CCAGGTGGCTGCTGACTTTCTTTAAAGAGCGGAGCGGGTGTGCGCAGACAGGGCG 2090
Db 1765 CCAGGTGGCTGCTGACTTTCTTTAAAGAGCGGAGCGGGTGTGCGCAGACAGGGCG 1824
QY 2091 CGCATTCATGATCATCCAGGAGCGGAGCCGGGCTGAGGAGTCTTCAAGCTCGAGCA 2150
Db 1825 CGCATTCATGATCATCCAGGAGCGGAGCCGGGCTGAGGAGTCTTCAAGCTCGAGCA 1884
QY 2151 GGATGTCTCCCAAGCAAGCTTCCACATCAGAGCTTGGGGTCAACAATCTTAACTG 2210
Db 1885 GGATGTCTCCCAAGCAAGCTTCCACATCAGAGCTTGGGGTCAACAATCTTAACTG 1944
QY 2211 CAGCCCAAGAGCGGAGAGAGCTAGACCTGTCTTCTGGTGACACTTACCCCAAGGAC 2270
Db 1945 CAGCCCAAGAGCGGAGAGAGCTAGACCTGTCTTCTGGTGACACTTACCCCAAGGAC 2004
QY 2271 TGTGAGCTTACGTGTCTCTCATCGAGCGGGTGGAGGTGGAGTCTTACTGCTGTGCG 2330
Db 2005 TGTGAGCTTACGTGTCTCTCATCGAGCGGGTGGAGGTGGAGTCTTACTGCTGTGCG 2064
QY 2331 CCTCGGGCTCATCTTGTGTGTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2390
Db 2065 CCTCGGGCTCATCTTGTGTGTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2124
QY 2391 TGTGGTATCTTACATGGCAATCATCTGATGATGCCGAGGCAAGCCAAAAAGTTTCA 2450
Db 2125 TGTGGTATCTTACATGGCAATCATCTGATGATGCC-AGGCAAGCCAAAAAGTTTCA 2183
QY 2451 GAAAGGGGAAAGAGCAATGATGCTCCATGTGTATGAGAGTCTTCAAGGACACCATGTGTA 2510
Db 2184 GAAAGGGGAAAGAGCAATGATGCTCCATGTGTATGAGAGTCTTCAAGGACACCATGTGTA 2243
QY 2511 TGGCATCTGCTTACAGAGATTCCAGCGGGCTCTTCTCTGAGCGCAGAGGTGACACTACCG 2570
Db 2244 TGGCATCTGCTTACAGAGATTCCAGCGGGCTCTTCTCTGAGCGCAGAGGTGACACTACCG 2303
QY 2571 GCCGTTCCAGGGCAACATGGGGGTGTCTCTCTCCCTCCACCAACCATATGCTCCAGGGC 2630
Db 2304 GCCGTTCCAGGGCAACATGGGGGTGTCTCTCTCCCTCCACCAACCATATGCTCCAGGGC 2363
QY 2631 CCCAAGTGCAAAGTTGGGCCACTGAGAGGCCACTCTCTGCTCCCTCTGAGTGTGAGAG 2690
Db 2364 CCCAAGTGCAAAGTTGGGCCACTGAGAGGCCACTCTCTGCTCCCTCTGAGTGTGAGAG 2423
QY 2691 TGAACGCTACACTTCTCCCATCCCAACAAATGGGATGTAAAGCAGCAAGACAGACAT 2750
Db 2424 TGAACGCTACACTTCTCCCATCCCAACAAATGGGATGTAAAGCAGCAAGACAGACAT 2483
QY 2751 TCCCTTACTGAACACTCAGAGGCCATGAGGCCAGCAGAAATTAATGATCCATTCCAGAC 2810
Db 2484 TCCCTTACTGAACACTCAGAGGCCATGAGGCCAGCAGAAATTAATGATCCATTCCAGAC 2543
QY 2811 GCTTGTGATGTTTCAATTAAGCAGGGGCACTGAGACCCCGCTGTTCTTCAACAGAA 2870
Db 2544 GCTTGTGATGTTTCAATTAAGCAGGGGCACTGAGACCCCGCTGTTCTTCAACAGAA 2603
QY 2871 TCTTAAGAGAGGAAATTAACAGAAAGAACAGAGAGGTTTCTGTGACACCGCCAAC 2930
Db 2604 TCTTAAGAGAGGAAATTAACAGAAAGAACAGAGAGGTTTCTGTGACACCGCCAAC 2663
QY 2931 TTTCAATTGCTCAGTGAATCTATTCTTAAGGCAAGACATTTGAAATGATGATTCACATC 2990
Db 2664 TTTCAATTGCTCAGTGAATCTATTCTTAAGGCAAGACATTTGAAATGATGATTCACATC 2723
QY 2991 TGGATACAGTATGACAGCTCATGCTGCTCACTTAAGGTGTGCGGTTGGCAGGCTG 3050
Db 2724 TGGATACAGTATGACAGCTCATGCTGCTCACTTAAGGTGTGCGGTTGGCAGGCTG 2783
QY 3051 TTAATGAGAGAGAGGCTGTGAGTCACTAGCATAGGGTTGACAGACCCCTGGATTGAG 3110
Db 2784 TTAATGAGAGAGAGGCTGTGAGTCACTAGCATAGGGTTGACAGACCCCTGGATTGAG 2843
QY 3111 AGTGTAAACAGAGGCTTGGCTTTCAGAGACAACTTCCAAATCCAGAGGCTTACT 3170
Db 2844 AGTGTAAACAGAGGCTTGGCTTTCAGAGACAACTTCCAAATCCAGAGGCTTACT 2903
QY 3171 GAGGTCCCTACTCTCACTGGGGTCCCGAGGTGAAACGACAAATGTGCTTTTATATT 3230
Db 2904 GAGGTCCCTACTCTCACTGGGGTCCCGAGGTGAAACGACAAATGTGCTTTTATATT 2963
QY 3231 ATTTATTTGGTGTCTGTGTATTATTAAGATCAAAATGTATAACCACTAGCTCTTTTC 3290
Db 2964 ATTTATTTGGTGTCTGTGTATTATTAAGATCAAAATGTATAACCACTAGCTCTTTTC 3023
QY 3291 ACCTGACTTGAATTAATCACTAATCTGATGATGCTTGGGTTGTGACTTCTACTG 3350
Db 3024 ACCTGACTTGAATTAATCACTAATCTGATGATGCTTGGGTTGTGACTTCTACTG 3083
QY 3351 ACCGCTGATTAACAGTGTGCTGCTGCCCCAGGTGGTGGGAATATTACAATCTGTCCA 3410
Db 3084 ACCGCTGATTAACAGTGTGCTGCTGCCCCAGGTGGTGGGAATATTACAATCTGTCCA 3143
QY 3411 CCAAGAAAGATGTGTGTGTGAGAGCAATGACATATCTGCTTGTGATTAAGAGACTT 3470
Db 3144 CCAAGAAAGATGTGTGTGTGAGAGCAATGACATATCTGCTTGTGATTAAGAGACTT 3203

QY	3471	CCGATTCCTCTAGAGTGGGTTCCGGTTATTCCTCATGTGTGGAAATTATCTTGAATCCATT	3530
Db	3204	CCGATTCCTCTAGAGTGGGTTCCGGTTATTCCTCATGTGTGGAAATTATCTTGAATCCATT	3263
QY	3531	GTCCATAGTCCTAGCAATTAAGAAATTTCCCAAGTTTCCATGTGGGGTTCCTCAGC	3590
Db	3264	GTCCATAGTCCTAGCAATTAAGAAATTTCCCTCAGATTTCATGTGGGGTTCCTCAGC	3323
QY	3591	TGCAGCAATCTTGGACATTTAAAGAAATTTAGAGAAATATTCATCCCTCTAAAATG	3650
Db	3324	TGCAGCAATCTTGGACATTTAAAGAAATTTAGAGAAATATTCATCCCTCTAAAATG	3383
QY	3651	TTTAAATATATACCAAAACAGTGGCCCCCTGCATTAAGTTTCTGTGGCCATGCACCAT	3710
Db	3384	TTTAAATATATACCAAAACAGTGGCCCCCTGCATTAAGTTTCTGTGGCCATGCACCAT	3443
QY	3711	TACTGGTACTTAAAAACAACACATTAAGTTATAGTCTGGGGATCAGAAATTCAAAAT	3770
Db	3444	TACTGGTACTTAAAAACAACACATTAAGTTATAGTCTGGGGATCAGAAATTCAAAAT	3503
QY	3771	GGATGTCCTGTAATGAAATCAAGGTGTGAGAGAGAGCTGTGCTCTTGTGAAGCTCAG	3830
Db	3504	GGATGTCCTGTAATGAAATCAAGGTGTGAGAGAGAGCTGTGCTCTTGTGAAGCTCAG	3553
QY	3831	GGAGAGCCGGTTCTTGGCATTTCAAGCTTCTAGAGCTGGCTCATTTCCAGGCTCA	3890
Db	3564	GGAGAGCCGGTTCTTGGCATTTCAAGCTTCTAGAGCTGGCTCATTTCCAGGCTCA	3623
QY	3891	GTGGCTGGTCAAGTTTCTCACAATGGCAATGCACTGACACAGGCGCCCTCCACTTCCCTC	3950
Db	3624	GTGGCTGGTCAAGTTTCTCACAATGGCAATGCACTGACACAGGCGCCCTCCACTTCCCTC	3683
QY	3951	TTTGACTTACAAAGCCCCCAGGAGAAATCCAGGATATCTCCATCTAAAGATCTTTCA	4010
Db	3684	TTTGACTTACAAAGCCCCCAGGAGAAATCCAGGATATCTCCATCTAAAGATCTTTCA	3743
QY	4011	TCATCTCTGGAAGAGCTTTTGGCATCACAAGACATAGCCACAGGTGGGATTAGAGCC	4070
Db	3744	TCATCTCTGGAAGAGCTTTTGGCATCACAAGACATAGCCACAGGTGGGATTAGAGCC	3803
QY	4071	AGGACATCTTTGGGGTGTGTTATTCGCTCCACACCTTCCTGCGCATACATCCACA	4130
Db	3804	AGGACATCTTTGGGGTGTGTTATTCGCTCCACACCTTCCTGCGCATACATCCACA	3863
QY	4131	GGAGAGCTCAAAAATATGATTTGGCGCACAGAGGATGTTTGTATGCTTGGCGACTCTAAC	4190
Db	3864	GGAGAGCTCAAAAATATGATTTGGCGCACAGAGGATGTTTGTATGCTTGGCGACTCTAAC	3923
QY	4191	ACCTTAAAAAACCACAGATCAGAAGATCTGGCATGTGGGCTCACATTTCTCACCTAGC	4250
Db	3924	ACCTTAAAAAACCACAGATCAGAAGATCTGGCATGTGGGCTCACATTTCTCACCTAGC	3983
QY	4231	AACAACGTGGCTGGAGCTGGGGACACAGCTGTGCTTTAAGAGGGGTGTCCACTTACACAGG	4310
Db	3984	AACAACGTGGCTGGAGCTGGGGACACAGCTGTGCTTTAAGAGGGGTGTCCACTTACACAGG	4043
QY	4311	TCACACAGCCACACATAGACGCTCTATCACCTCCACAAATGAGAGCTAAGGTTTGTGTTCTA	4370
Db	4044	TCACACAGCCACACATAGACGCTCTATCACCTCCACAAATGAGAGCTAAGGTTTGTGTTCTA	4103
QY	4371	CTGATCAATGCCCCCTGCGAGGTTCATTTATGTAAAGAAAAAGAAAGACTGGGATTAATC	4430
Db	4104	CTGATCAATGCCCCCTGCGAGGTTCATTTATGTAAAGAAAAAGAAAGACTGGGATTAATC	4153
QY	4431	TCCTATCAGGTGAGATAGACATAGACCAATGTGTGTACACTTACCTTTTCTTTTCTTTT	4490
Db	4164	TCCTATCAGGTGAGATAGACATAGACCAATGTGTGTCTACACTTACCTTTTCTTTTCTTTT	4223
QY	4491	TTCTTTTCTTTTCTTTTCTTTTAAATGAGAGAGATCTCATTTCTGTGCTTGAAGC	4550
Db	4224	TTCTTTTCTTTTCTTTTCTTTTAAATGAGAGAGATCTCATTTCTGTGCTTGAAGC	4283
QY	4551	TGGAGTCAGTGGCGCAANTTCGGCTCACTGCAACCTCTGCCCTCCAGGACTTCAAGCAATT	4610

Db	4284	TTGGAGTSCAGTGGGGGCAATGCTGAGCTCAGTCCGCAACCTCTGCGGCTCAAGCAATT	4343
Oy	4611	CTCCACCTCAGCCTCTCCCAATATAGCTGGGATCACTGGCCAAACACCATCGCCAGCTAA	4670
Db	4344	CTCCCACTCAGCCTCCCAATATAGCTGGGATCACTGGGCAAAACCAACCATCGCCAGCTAA	4403
Oy	4671	TTTTGTATTTTTGTAGAGACAGGGTTTCACCATGTGTGCCAGGCTGGTCTCAACCTCT	4730
Db	4404	TTTTGTATTTTTGTAGAGACAGGGTTTCACCATGTGTGCCAGGCTGGTCTCAACCTCT	4463
Oy	4731	GGGCTCAAGCAATCTCTCTGCTGGGCTTCCCAAAGTGTGGGATTAACAGATGTGAGCA	4780
Db	4464	GGGCTCAAGCAATCTCTCTGCTGGGCTTCCCAAAGTGTGGGATTAACAGATGTGAGCA	4523
Oy	4791	CCGATCCAGGCCCCACACCCCTCATTTATACCAATTAACGCCCCAGTAACCTGGGCACTTTT	4850
Db	4534	CCGATCCAGGCCCCACACCCCTCATTTATACCAATTAACGCCCCAGTAACCTGGGCACTTTT	4583
Oy	4851	GCTTCTCACCCCTGCTCTGATCTGGAAGAGAGGGATTAATGTATAGCTTGTGACACA	4910
Db	4584	GCTTCTCACCCCTGCTCTGATCTGGAAGAGAGGGATTAATGTATAGCTTGTGACACA	4643
Oy	4911	GTCCCAAGTCAATATTTCTGCGGCAAAAATTCTTCAAAAAATAATGTACTTCAATG	4970
Db	4644	GTCCCAAGTCAATATTTCTGCGGCAAAAATTCTTCAAAAAATAATGTACTTCAATG	4703
Oy	4971	TATTCATGAAATTCACCTTGGAAATGACCGGCTCAACTGTGTCAATGGCAATAATGAA	5030
Db	4704	TATTCATGAAATTCACCTTGGAAATGACCGGCTCAACTGTGTCAATGGCAATAATGAA	4763
Oy	5031	AGGAATTTATATAGTCTCTTAATGGGCTGTACTGCAAGACCTCTTGAACACTTCCAGAG	5090
Db	4764	AGGAATTTATATAGTCTCTTAATGGGCTGTACTGCAAGACCTCTTGAACACTTCCAGAG	4823
Oy	5091	GATAGGATATTTAAGTATATGCTATGCGCTTGGCGTATGGCACTTCCCTTGTAAATG	5150
Db	4824	GATAGGATATTTAAGTATATGCTATGCGCTTGGCGTATGGCACTTCCCTTGTAAATG	4883
Oy	5151	TGGTTCCGGCCCACTGACCCCTTGGCGCTGTGAGCCAGATGCTGACCCGATTAAGGAGC	5210
Db	4884	TGGTTCCGGCCCACTGACCCCTTGGCGCTGTGAGCCAGATGCTGACCCGATTAAGGAGC	4943
Oy	5211	CAAAAGAGGGCTGCGGCTTCTTCCCTCACTGTAAGACCCCTTATTTGAATTCACGTGTG	5270
Db	4944	CAAAAGAGGGCTGCGGCTTCTTCCCTCACTGTAAGACCCCTTATTTGAATTCACGTGTG	5003
Oy	5271	GAGCCCTAGCCCTCATTTCTGCAATYCCCAACCTCCAGCCCTTCCAMGACGAGACTA	5330
Db	5004	GAGCCCTAGCCCTCATTTCTGCAATYCCCAACCTCCAGCCCTTCCAMGACGAGACTA	5063
Oy	5331	GSTGCCCTGCATTCACCCAAAGGTGGGATTTGGGCTTCCTTAAGGCGGCTACTGTGCACA	5390
Db	5064	GSTGCCCTGCATTCACCCAAAGGTGGGATTTGGGCTTCCTTAAGGCGGCTACTGTGCACA	5133
Oy	5391	TCACCGACATCATCTGTGCTGTGCAGAGACACACGCGGCATTTTCTTCAACTAGAGGC	5450
Db	5124	TCACCGACATCATCTGTGCTGTGCAGAGACACACGCGGCATTTTCTTCAACTAGAGGC	5183
Oy	5451	TCAAAACCTCTGACAGATTTGCTGGCTCTGAGACACGATTTTCCGAGCTGTGCTCA	5510
Db	5184	TCAAAACCTCTGACAGATTTGCTGGCTCTGAGACACGATTTTCCGAGGTMGTCTCA	5243
Oy	5511	GTGAAGGGGGCCACCTCGAGGAACCCCTGGCTCTTTCTTAAAGCCACAGCCCACTTAC	5570
Db	5244	GTGAAGGGGGCCACCTCGAGGAACCCCTGGCTCTTTCTTAAAGCCACAGCCCACTTAC	5303
Oy	5571	ATAAAAACATTTCAAGGCTCAGTGGAAACAGTGAAGTCCATTTGTGTGAAGCTACTGATG	5630
Db	5304	ATAAAAACATTTCAAGGCTCAGTGGAAACAGTGAAGTCCATTTGTGTGAAGCTACTGATG	5631
Oy	5631	CCAGCCACATGCTCATTCACAGTGGTCTGCCATGCCCTACAGGAAGGCCAGCGCATGC	

Db	5364	CCAGCCCACTGCTCATCTACAGTGGTATGACCATGCTACAGAGGAAGCCAGCCCATGCAAG	5423
Oy	5691	ACTGGTCTCTAAATCTCTGTGCTATTGTCACAGAAAGGAAAGTCTCAAGAAAGTCAACT	5750
Db	5424	ANTGGTCTCTAAATGNTTGTGTCTATTGTCACAGAAAGGAAAGTCTCAAGAAAGTCAACT	5483
Oy	5751	GGGACAAAGCACAAAGCCCAAGCGGACATAGGCGCTTGTAAGAGTTAGCAGACTGGTGTGTG	5810
Db	5484	GGGACAAAGCACAAAGCCCAAGCGGACATAGGCGCTTGTAAGAGTTAGCAGACTGGTGTGTG	5543
Oy	5811	GATCTGCAGTCTTCACTGGAATTAATTATTCATTGCAAGATCTTTTAAAGTGGCAATT	5870
Db	5544	GATCTGCAGTCTTCACTGGAATTAATTATTCATTGCAAGATCTTTTAAAGTGGCAATT	5603
Oy	5871	TATTCATTTCCGTGCTTTAAATAAACAATGACCAAAACAAAGTATCAAGCTGTTA	5930
Db	5604	TATTCATTTCCGTGCTTTAAATAAACAATGACCAAAACAAAGTATCAAGCTGTTA	5663
Oy	5931	AGTCGTTCCGGCTACTTGTCCCTGGTTCAGATAGAGCCCCGGTTTCCAGTTTGTGACTG	5990
Db	5664	AGTCGTTCCGGCTACTTGTCCCTGGTTCAGATAGAGCCCCGGTTTCCAGTTTGTGACTG	5723
Oy	5991	TGACAGGCTCAGCATGGGCTCAGCAGATGGTGCTTAATTGTGGATGATACAGAAACC	6050
Db	5724	TGACAGGCTCAGCATGGGCTCAGCAGATGGTGCTTAATTGTGGATGATACAGAAACC	5783
Oy	6051	AGGCTTTGGGATACAAAGTCTTTCTCTTCATTGTGATGGCGTGCAGCTGTGTGAACAGAT	6110
Db	5784	AGGCTTTGGGATACAAAGTCTTTCTCTTCATTGTGATGGCGTGCAGCTGTGTGAACAGAT	5843
Oy	6111	GTTTTTCGCGGAATTAATAAATAATAGTCTGGAGCTCTGCCAAAAAATAAAAAA 6163	
Db	5844	GTTTTTCGCGGAATTAATAAATAATAGTCTGGAGCTCTGCCAAAAAATAAAAAA 5896	

XX	RESULT 3
XX	AAH18243
ID	AAH18243 standard; cDNA; 5573 BP.
XX	
AC	AAH18243;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:18190.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES. INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
XX	

xx Claim 8, SEQ ID 18190; 2537pp + CD ROM, English.

xx The present invention describes primer sets for syn

cc full-length cDNAs defined in the specification. Whe

cc comprises: (a) an oligo-dT primer and an oligonucle

cc to the complementary strand of a polynucleotide whi

cc the 5602 nucleotide sequences defined in the specif

cc oligonucleotide comprises at least 15 nucleotides; or

cc of an oligonucleotide comprising a sequence complementary to the

cc complementary strand of a polynucleotide which comprises a 5'-end

cc sequence and an oligonucleotide comprising a sequence complementary to a

cc polynucleotide which comprises a 3'-end sequence, where the

cc oligonucleotide comprises at least 15 nucleotides and the combination of

cc the 5'-end sequence/3'-end sequence is selected from those defined in

cc the specification. The primer sets can be used in antisense therapy and

cc in gene therapy. The primers are useful for synthesizing polynucleotides

cc particularly full-length cDNAs. The primers are also useful for the

cc detection and/or diagnosis of the abnormality of the proteins encoded by

cc the full-length cDNAs. The primers allow obtaining of the full-length

cc cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

cc AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

cc AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

cc represent oligonucleotides, all of which are used in the exemplification

cc of the present invention.

xx

xx

xx Sequence 5573 BP; 1373 A; 1524 C; 1290 G; 1386 T; 0 other;

[illegible]

Dh 600 AGCTGGAGCAAGACGCTGGGAAACATGCGGGGAACCTTCAACTCTCTCTGCAAGGCT 659
Qy 1229 GTGACCAAGATGCCAAAGATCCAGGGATCTCGGCTGAGTTCGAAGTTTGGTCCAC 1288
Dh 660 GTGACCAAGATGCCAAAGATCCAGGGATCTCGGCTGAGTTCGAAGTTTGGTCCAC 719
Qy 1289 ATCCACAAATGTAAGCAATAAATCTAGCTGGTGTAGTATGAGCAGCATGT 1348
Dh 720 ATCCACAAATGTAAGCAATAAATCTAGCTGGTGTAGTATGAGCAGCATGT 779
Qy 1349 CACTCACCATCGAGCCGCGGTCAAAACAGAGCCGAAGTTTGTCCCTGGCTTTTCG 1408
Dh 780 CACTCACCATCGAGCCGCGGTCAAAACAGAGCCGAAGTTTGTCCCTGGCTTTTCG 839
Qy 1409 TGTGTCTAATCTGGGAGCTGAGTGAAGACCTCAGCCTGACATGTGGCTCCAAACAC 1468
Dh 840 TGTGTCTAATCTGGGAGCTGAGTGAAGACCTCAGCCTGACATGTGGCTCCAAACAC 899
Qy 1469 AATCTCTCTCTTGTGATCTGACACGTCTGTGATGTAATGTGAAAAACATTA 1528
Dh 900 AATCTCTCTCTTGTGATCTGACACGTCTGTGATGTAATGTGAAAAACATTA 959
Qy 1529 GCTGCACAGACCAACGGTACTGCAAAAGAAATCTACTACTCAGAGTCCAGTGA 1588
Dh 960 GCTGCACAGACCAACGGTACTGCAAAAGAAATCTACTACTCAGAGTCCAGTGA 1019
Qy 1589 TCTCTCAGCTGCTGTGAGAGCTGATGACTTCTCTGGAAGCTGTGTGCTCCAAAGACA 1648
Dh 1020 TCTCTCAGCTGCTGTGAGAGCTGATGACTTCTCTGGAAGCTGTGTGCTCCAAAGACA 1079
Qy 1649 GGCTCAGCTGTGTGTGTGAGAGCTGCAAGAGCTGAGAGCATACACAGAAAGCCCT 1708
Dh 1080 GGCTCAGCTGTGTGTGTGAGAGCTGCAAGAGCTGAGAGCATACACAGAAAGCCCT 1139
Qy 1709 GCAACACAGCTTCACTACTCTGCGCAGAGTCCATCCACAGCAGAGCTTACTTGC 1768
Dh 1140 GCAACACAGCTTCACTACTCTGCGCAGAGTCCATCCACAGCAGAGCTTACTTGC 1199
Qy 1769 GCTCTCTTGTGCGGAGAGCTCTATCAGACATCCAGGTGAAGCAGAAATCTGCTGA 1828
Dh 1200 GCTCTCTTGTGCGGAGAGCTCTATCAGACATCCAGGTGAAGCAGAAATCTGCTGA 1259
Qy 1829 CCCTTGCACCTTGTGCGGAGAGCTTCAACAGAGGCTTCAGAGGCTGAGGCTGAGGCTG 1888
Dh 1260 CCCTTGCACCTTGTGCGGAGAGCTTCAACAGAGGCTTCAGAGGCTGAGGCTGAGGCTG 1319
Qy 1889 CCTTATACCTTATTTCAAGAGAGAGGCTTTTCAAGGTGAGCCCTGACACAAAAAGCA 1948
Dh 1320 CCTTATACCTTATTTCAAGAGAGAGGCTTTTCAAGGTGAGCCCTGACACAAAAAGCA 1379
Qy 1949 AGGTCTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCTCTGAGCTGTGCT 2008
Dh 1380 AGGTCTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCTCTGAGCTGTGCT 1439
Qy 2009 GGAACATCAGCGGTGCGCAGAGACAGGTGGCTGCTGACTTCTTTAAGAGAGGAGG 2068
Dh 1440 GGAACATCAGCGGTGCGCAGAGACAGGTGGCTGCTGACTTCTTTAAGAGAGGAGG 1499
Qy 2069 GGTGTGTGCGCAGAGAGGCGGCTTATGATATCTCAGAGAGCAGCGAGCCGGGCTG 2128
Dh 1500 GGTGTGTGCGCAGAGAGGCGGCTTATGATATCTCAGAGAGCAGCGAGCCGGGCTG 1559
Qy 2129 AGGAGATCTTCAAGCTTGAAGAGAGATGTGCTCCCAAGCCAGCTTCCACATCAGCT 2188
Dh 1560 AGGAGATCTTCAAGCTTGAAGAGAGATGTGCTCCCAAGCCAGCTTCCACATCAGCT 1619
Qy 2189 TGTGGGTCAACATCTTAATCTGAGCCCAACAGAGGGGCAAGCAGCTAGACTCTTCT 2248
Dh 1620 TGTGGGTCAACATCTTAATCTGAGCCCAACAGAGGGGCAAGCAGCTAGACTCTTCT 1679
Qy 2249 CGGTGACACTTACCCCAAGAGAGTGTGAGTGTGACTGTCACTCTCATCGCAGCGGTGGAG 2308
Dh 1680 CGGTGACACTTACCCCAAGAGAGTGTGAGTGTGACTGTCACTCTCATCGCAGCGGTGGAG 1739

Qy 2309 GTGAGAGCTTACGTGTGTGCGGCTGACATCTTGTGTGAAAAAGAAAA 2368
Dh 1740 GTGAGAGCTTACGTGTGTGCGGCTGACATCTTGTGTGAAAAAGAAAA 1799
Qy 2369 AGAAGACAAAGAGGCGCGCTGTGAGTATCTACATGGCAACATCAATCTAGATGC 2428
Dh 1800 AGAAGACAAAGAGGCGCGCGCTGTGAGTATCTACATGGCAACATCAATCTAGATGC 1859
Qy 2429 CGAGGACGCAAAAAAGTTTCAGAAAAGGCGAAAGCAATGACTCCATGTGTATGAC 2488
Dh 1860 CGAGGACGCAAAAAAGTTTCAGAAAAGGCGAAAGCAATGACTCCATGTGTATGAC 1919
Qy 2489 TCAATGAGGACACATGTGTATGGGATGTGTACAGGATTCAGAGGCTGCTTCTG 2548
Dh 1920 TCAATGAGGACACATGTGTATGGGATGTGTACAGGATTCAGAGGCTGCTTCTG 1979
Qy 2549 AGCCAGAGGTGACACCTTACCGGCGCTTCAGAGGACACATGGGGCTGTCTCCCTCC 2608
Dh 1980 AGCCAGAGGTGACACCTTACCGGCGCTTCAGAGGACACATGGGGCTGTCTCCCTCC 2039
Qy 2609 CACCCACATATGCTCCAGGGCGCCCAACTGCAAAAGTTGGCCACTGAGAGCCACTCTC 2668
Dh 2040 CACCCACATATGCTCCAGGGCGCCCAACTGCAAAAGTTGGCCACTGAGAGCCACTCTC 2099
Qy 2669 GCTCCCTCTGAGTGTGAGAGTGAACCCGTACACCTTCTCCATCCCAATAGGGATG 2728
Dh 2100 GCTCCCTCTGAGTGTGAGAGTGAACCCGTACACCTTCTCCATCCCAATAGGGATG 2159
Qy 2729 TAAGCAGAGGACACAGACATTCCTTACTGAACACTGAGAGCCATGAGCCAGAC 2788
Dh 2160 TAAGCAGAGGACACAGACATTCCTTACTGAACACTGAGAGCCATGAGCCAGAC 2219
Qy 2789 AATTAATGATTCACAGCGCTTGTGATGATTAAGCAGGCACTGAGACACC 2848
Dh 2220 AATTAATGATTCACAGCGCTTGTGATGATTAAGCAGGCACTGAGACACC 2279
Qy 2849 CGTCCGCTGTCTTACCAAGAAATCTTAAGAAAGAAATTAACAGAAAGAACAGAGA 2908
Dh 2280 CGTCCGCTGTCTTACCAAGAAATCTTAAGAAAGAAATTAACAGAAAGAACAGAGA 2339
Qy 2909 GGTTCCTGTGACACGCGCACTTCACTGTCAGTGAAGTCAATTTAAGGCAAGACA 2968
Dh 2340 GGTTCCTGTGACACGCGCACTTCACTGTCAGTGAAGTCAATTTAAGGCAAGACA 2399
Qy 2969 TTGAAAATGATGAATTCGAATCTGATACAGTATGACAGCTCATGTGCTCAACTTA 3028
Dh 2400 TTGAAAATGATGAATTCGAATCTGATACAGTATGACAGCTCATGTGCTCAACTTA 2459
Qy 3029 GGTGTGCGGTATGACCAAGCTGTATGAGAGAGAGAGGCTGAGTACCTATAGG 3088
Dh 2460 GGTGTGCGGTATGACCAAGCTGTATGAGAGAGAGAGGCTGAGTACCTATAGG 2519
Qy 3089 TTGCAGAGCCCTGAGTTCAGAGTGTAAACAGAGGCTTGGCCCTTCAGAGCAACGT 3148
Dh 2520 TTGCAGAGCCCTGAGTTCAGAGTGTAAACAGAGGCTTGGCCCTTCAGAGCAACGT 2579
Qy 3149 TCCAAATTCAGAGGCTTACAGGCTCCTACTCTCAGTGGGCTCCAGATGAAAC 3208
Dh 2580 TCCAAATTCAGAGGCTTACAGGCTCCTACTCTCAGTGGGCTCCAGATGAAAC 2639
Qy 3209 GACAATGTGCTTTTATTTATTTATTTATTTGAGTGTCTGTATTTAAGATCAAT 3268
Dh 2640 GACAATGTGCTTTTATTTATTTATTTATTTGAGTGTCTGTATTTAAGATCAAT 2699
Qy 3269 GTATTAACACCTTACCTTTTCACTGACTTATTAATTAATCAATCAATCAATGTTGGAT 3328
Dh 2700 GTATTAACACCTTACCTTTTCACTGACTTATTAATTAATCAATCAATCAATGTTGGAT 2759
Qy 3329 GCTGTGGTGTGACTTCTACTGACCGCTAGATTAACGTGTGCTGTCCAGAGTGGTG 3388
Dh 2760 GCTGTGGTGTGACTTCTACTGACCGCTAGATTAACGTGTGCTGTCCAGAGTGGTG 2819

QY 655 TTCATCTGGAGATGTCAAAGCTCATAGAGCATGGTTTAAAGCTGAGTTTTCATCCCT 714
|||||
Db 484 TTCATCTGGAGATGTCAAAGCTCATAGAGCATGGTTTAAAGCTGAGTTTTCATCCCT 543
QY 715 CGCCTGAGGAGATGCGGTCCGGGTGAGAGCTGCCACAGAGTCACTCACTCATCAGC 774
|||||
Db 544 CGCATGAGGAGATGCGGTCCGGGTGAGAGCTGCCACAGAGTCACTCACTCATCAGC 603
QY 775 GGCAGATTCAGATGCCACCGTGTGAGAGTGGAACTTTCTGCAGCAATGGCACTGTCTC 834
|||||
Db 604 GGCAGATTCAGATGCCACCGTGTGAGAGTGGAACTTTCTGCAGCAATGGCACTGTCTC 663
QY 835 CGGAT -CAAGATGCAAGAGAGTGAATGGCTTACACCTCCCATGG -TTCCACCCC 891
|||||
Db 664 CGGATTCAGATGCCACCGTGTGAGAGTGGAACTTTCTGCAGCAATGGCACTGTCTC 723
QY 892 AGAAATGCTCCGCTTACAGATTCGCAAAACCGCTCATCTATTAACACTGCTGATCATC 951
|||||
Db 724 GAAATGCTCCGCTTACAGATTCGCAAAACCGCTCATCTATTAACACTGCTGATCATC 783
QY 952 GAGTCTGTGTTGAGGTGGAAGGCTCAGCAACCTG -ATGCTGCCAATCAACCA -GAAG 1009
|||||
Db 784 GAGTCTGTGTTGAGGTGGAAGGCTCAGCAACCTGGAATGCTGTCCCAATCCACGAG 843
QY 1010 GCTTCCCTGAGATGAGCTCATGACGTGGCAGTTTGTCTTGCACACCTGCGGCCA 1069
|||||
Db 844 GCTTCCCTGAGATGAGCTCATGACGTGGCAGTTTGTCTTGCACACCTGCGGCCA 903
QY 1070 GCGTCTCTTCTCAACTTCACTCACTCACTCACTGAGAGAGAGAGAGAGAGAGAGAG 1129
|||||
Db 904 GCGTCTCTTCTCAACTTCACTCACTCACTCACTGAGAGAGAGAGAGAGAGAGAGAG 963
QY 1130 ACTACATCCC -GGGCTCCACCA -CCAACCCCGAGAGTGTTCAGCTGAGAGCAAGCAGCC 1187
|||||
Db 964 ACTACATCCCAGGCTCCACACCCCAACCCCGAGAGTGTTCAGCTGAGAGCAAGCAGCC 1023
QY 1188 TGGGACATGCGGGGGAACCTTCAACT -CTCTGCAAGGCTGTGACCAAGATG -CCCA 1244
|||||
Db 1024 TGGGACATGCGGGGGAACCTTCAACTCTCTGCGCAAAAGCTGTGACCAAGATGCCCA 1083
QY 1245 AAGTCCAGGGATCT -CCGGCTGAGTTCAGAGTTTGGTCCCAACCAAAATG -AA 1302
|||||
Db 1084 AAGTCCAGGGATCTCCGGCTGAGTTCAGAGTTTGGTCCCAACCAAAATG -AA 1143
QY 1303 AGCAATTAATCTACGTGTGACTGTGATGAGAGGAGCATGTCACTCAGCATGAG 1362
|||||
Db 1144 AGCAATTAATCTACGTGTGACTGTGATGAGAGGAGCATGTCACTCAGCATGAG 1203
QY 1363 CCACGCGCCGTCAAAACAGAGCCGCAAGTTTGTCCGTGTGTCTGATGAATCT 1422
|||||
Db 1204 CCATGCGCCGTCAAAACAGAGCCGCAAGTTTGTCCGTGTGTCTGATGAATCT 1263
QY 1423 CGGACCTGAGTGAACATCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1482
|||||
Db 1264 CGGACCTGAGTGAACATCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1323
QY 1483 TGTGATGATCTGACACGTCTGTGGATGAATGGAAAAAACATTAAGCTCAGACAGCAC 1542
|||||
Db 1324 TGTGATGATCTGACACGTCTGTGGATGAATGGAAAAAACATTAAGCTCAGACAGCAC 1383
QY 1543 CGGATCTGCAAAAGAAATCTTA -CTCAGTCCAGG -TGCCAGTGAATCTTCCACCTG 1599
|||||
Db 1384 CGGATCTGCAAAAGAAATCTTAAGTCTCACTCACTCACTCACTCACTCACTCACTG 1443
QY 1600 CCTGTGAGAGTGAATGACTTCT -CCTGGAAGCTGCTGG -TGCCCAAGAGACAGG -TGAGC 1656
|||||
Db 1444 CCTGTGAGAGTGAATGACTTCTCCTGTGGAAGCTGTGGATGAGCCAGGAGAGGCTTAGC 1503
QY 1657 CTGGT -GCTGTGAGAGAGAGAGTGTGAGAGAGATACAGAGAGAGAGAGAGAGAGAG 1715
|||||
Db 1504 CTGGTGTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1563

QY 1716 CAGCTTACGTACCTCTGCGCAGTCCATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1775
|||||
Db 1564 CAGCTTACGTACCTCTGCGCAGTCCATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1623
QY 1776 CTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1832
|||||
Db 1624 CTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1683
QY 1833 TCACA -CCTTGTGCCCCAGCTTCCA -ACAAGAGAGCTCCAGAGAGAGAGAGAGAGAGAG 1887
|||||
Db 1684 TCAGACCTTTTGGCCAGAGCTTCATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1743
QY 1888 TCCCTTATACCTTATTTCAAG -AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1945
|||||
Db 1744 TCCCTTATACCTTATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
QY 1946 GCAG -GGTCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2004
|||||
Db 1804 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1863
QY 2005 TCCGTGAGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2064
|||||
Db 1864 TCCGTGAGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1923
QY 2065 AGCGGCTGTGTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124
|||||
Db 1924 AGCGGCTGTGTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1983
QY 2125 GCTGAGAGAGATCTTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2184
|||||
Db 1984 GCTGAGAGAGATCTTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2043
QY 2185 AGCTTCTGGGTCAACATCTCTTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2244
|||||
Db 2044 AGCTTCTGGGTCAACATCTCTTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2103
QY 2245 TTCTCGGTGACACTTACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
|||||
Db 2104 TTCTCGGTGACACTTACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2163
QY 2305 GAGAGTGAAGCTTACAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2363
|||||
Db 2164 GAGAGTGAAGCTTACAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2223
QY 2364 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2423
|||||
Db 2224 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2283
QY 2424 GATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2483
|||||
Db 2284 GATGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2343
QY 2484 TGCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2539
|||||
Db 2344 TGCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2403
QY 2540 CTTTCTGTCAG -CCAGAGTGTGACACTTAC -CGGCGCTTCCAGAGAG -ACCATGGGGTCT 2596
|||||
Db 2404 CTTTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463
QY 2597 GTCTCTCC -TCCGACCCACATATGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2652
|||||
Db 2464 GTCTCTCCGTCCTCCACCATATGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
QY 2653 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2712
|||||
Db 2524 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583
QY 2713 CCCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2771
|||||
Db 2584 CCCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2643
QY 2772 GCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2830

```

Db 2644 GCCCATGAGCCAGCAGAACTATGATCCATTTGACAGCGCTTGTGAGTTTATATA 2703
Oy 2831 CGAGGGCACTGAGACACCCGCTCCGTCTTCTTACCAAGAAATCTTAAAGAGGAATAT 2890
Db 2704 CGAGGGCGCTGAGACACCCGCTCCGTCTTCTTACCAAGAAATCTTAAAGAGGAATAT 2763
Oy 2891 ACAGAGGAACAGCAGAGGATTTCTGTG-GACACCGCCCAATCTCATATTCAGTGGAC 2949
Db 2764 ACAGAGGAACAGCAGAGGATTTCTGTGAGACACCGCCCAATCTCATATTCAGTGGAC 2823
Oy 2950 TCATTCTAAGGGCAAGACATTTGAAATGATGATTCATTCATTCATTCATTCATTCATTC 3009
Db 2824 TCATTCTAAGGGCAAGACATTTGAAATGATGATTCATTCATTCATTCATTCATTCATTC 2883
Oy 3010 TCATGTGCTCTCAACTTA-----GGCTGTGGGTTAGCCAGCCCTGT---AATGAGAGA 3061
Db 2884 TCATGTGCTCTCAACTTAAGCAATGAGCATGTGACAGTTTACCAAGCCATGTAATGAGAGA 2943
Oy 3062 GAGAGGCGCTGAGTCACTAGCATAGGTTGACAGCAAGCCCTGATTCAGAGTTTAAACA 3121
Db 2944 GAGAGGCGCTGAGTCACTAGCATAGGTTGACAGCAAGCCCTGATTCAGAGTTTAAACA 3003
Oy 3122 GAGGCGTCCCTCTTCCAGACAAAGTTCCAAATTCAGAGGAGCTTACCTGAGTCCCTAC 3181
Db 3004 GAGGCGTCCCTCTTCCAGACAAAGTTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3063
Oy 3182 TCTCACTGGGGTCCCGAGGATGAAACGACAAATGTCCTTTTATTTATTTATTTATTTATTT 3241
Db 3064 TCTCACTGGGGTCCCGAGGATGAAACGACAAATGTCCTTTTATTTATTTATTTATTTATTT 3123
Oy 3242 GGTCTGTGTATTTAAGAGA---TCAAAATGATTAACCACTAGGCTTTTACCGTACT 3298
Db 3124 GGTCTGTGTATTTAAGAGAATCTGAGTGTATTAACCACTAGGCTTTTACCGTACT 3183
Oy 3299 TAGTATTAATCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3356
Db 3184 TAGTATTAATCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3243
Oy 3357 AGATTAAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 3414
Db 3244 AGATTAAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 3301

```

RESULT 5

AAA26351

ID AAA26351 standard: cDNA, 2209 BP.

AC AAA26351;

XX 29-JUN-2000 (first entry)

XX Human secreted protein gene 6 SEQ ID NO:16.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

XX antiHIV; antiinflammatory; neurotrophic; neuroprotective; antiallergic;

XX osteoporotic; antiarthritic; antibacterial; antidiabetic; antiasthma;

XX antiproliferic; cardiant; gene therapy; cancer; neurological disorder;

XX immune disease; inflammation; blood disorder; tumor; ss.

XX Homo sapiens.

XX WO200006698-A1.

XX 10-FEB-2000.

XX 29-JUL-1999; 99WO-US17130.

XX 30-JUL-1998; 98US-0094657.

XX 05-AUG-1998; 98US-0095486.

XX 06-AUG-1998; 98US-0095454.

XX 12-AUG-1998; 98US-0095455.

```

XX (HUMA-) HUMAN GENOME SCI INC.
PA Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
XX Latleir D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Edress GA, Edner R, Olsen HS, Mucenski M;
XX WPI: 2000-195282/17.
DR P-PSDB: AA91456.
XX New isolated human genes and the secreted polypeptides they encode,
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 375; 634pp; English.
XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
XX human secreted proteins given in AA91451 to AA91491. The human secreted
XX proteins can have activities based on the tissues and cells they are
XX expressed in. Examples of the activities are: cytostatic;
XX immunosuppressive; antiHIV; antiinflammatory; neurotrophic; neuroprotective;
XX antiallergic; osteoporotic; antiarthritic; antibacterial; antidiabetic;
XX antiasthma; antiproliferic; and cardiant. The polynucleotides and their
XX corresponding secreted proteins are useful for preventing, treating or
XX ameliorating medical conditions, e.g. by protein or gene therapy. Also
XX pathological conditions can be diagnosed by determining the amount of the
XX proteins in a sample or by determining the presence of mutations in the
XX polynucleotides. Specific uses are described for each of the
XX polynucleotides, based on which tissues they are most highly expressed
XX in, and include developing products for the diagnosis or treatment of
XX cancer, tumours, neurodegenerative disorders, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, inflammation,
XX allergies, Alzheimer's and behavioural disorders, schizophrenia,
XX osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
XX transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
XX cardiovascular disorders, reproductive disorders, gastrointestinal
XX disorders, respiratory disorders and metabolic disorders. The proteins
XX or polynucleotides can also be used as food additives or preservatives.
XX The proteins are also useful for identifying their binding partners.
XX AAA26337 to AAA26345 and AA91450 are sequences used in the
XX exemplification of the present invention.
SQ Sequence 2209 BP; 508 A; 646 C; 566 G; 477 T; 12 other:
Query Match 35.4%; Score 2179; DB 21; Length 2209;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 2189; Conservative 12; Mismatches 7; Indels 3; Gaps 1;
Oy 204 AGCTGGCGGGGCTTGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 263
Db 2 AGCTGGCGGGGCTTGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 61
Oy 264 GCCGAGGCGTCCCGAGGATGAGGCGGCTGAACTCGGGGCTCTATGCACTGTAGG 323
Db 62 GCCGAGGCGTCCCGAGGATGAGGCGGCTGAACTCGGGGCTCTATGCACTGTAGG 121
Oy 324 GGTTCGTGCTGCTGGTGGCGGCGGCTGCGCGCGGGGCGAGAAAGCTTTTGAATGCTCT 383
Db 122 GGTTCGTGCTGCTGGTGGCGGCGGCTGCGCGCGGGGCGAGAAAGCTTTTGAATGCTCTCT 181
Oy 384 GCCACGAGAAAGCAACATTAAGTTCTCAATTAAGCTGGGGAGCCCGACCTCTGGCAAA 443
Db 182 GCCACGAGAAAGCAACATTAAGTTCTCAATTAAGCTGGGGAGCCCGACCTCTGGCAAA 241
Oy 444 ACCCTGTACATCGTCAATTTCTAAAGACATTAACCATGTTGTCATGAAGTCTGAGA 503
Db 242 ACCCTGTACATCGTCAATTTCTAAAGACATTAACCATGTTGTCATGAAGTCTGAGA 301
Oy 504 AAGATATGCTTTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 563
Db 302 AAGATATGCTTTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 361

```


PF 01-DEC-2000; 2000MO-US32678.
 XX
 PR 01-DEC-1999; 99MO-US28301.
 PR 01-DEC-1999; 99MO-US28634.
 PR 02-DEC-1999; 99MO-US28531.
 PR 02-DEC-1999; 99MO-US28564.
 PR 02-DEC-1999; 99MO-US28565.
 PR 09-DEC-1999; 99MO-US170262.
 PR 16-DEC-1999; 99MO-US30095.
 PR 20-DEC-1999; 99MO-US30911.
 PR 20-DEC-1999; 99MO-US30911.
 PR 30-DEC-1999; 99MO-US31273.
 PR 06-JAN-2000; 2000MO-US002247.
 PR 11-FEB-2000; 2000MO-US00376.
 PR 11-FEB-2000; 2000MO-US03565.
 PR 18-FEB-2000; 2000MO-US04341.
 PR 18-FEB-2000; 2000MO-US04342.
 PR 22-FEB-2000; 2000MO-US04914.
 PR 24-FEB-2000; 2000MO-US04914.
 PR 24-FEB-2000; 2000MO-US05004.
 PR 01-MAR-2000; 2000MO-US05601.
 PR 20-MAR-2000; 2000MO-US07377.
 PR 21-MAR-2000; 2000MO-US07532.
 PR 30-MAR-2000; 2000MO-US08439.
 PR 17-MAY-2000; 2000MO-US13705.
 PR 22-MAY-2000; 2000MO-US14042.
 PR 30-MAY-2000; 2000MO-US14941.
 PR 02-JUN-2000; 2000MO-US15264.
 PR 10-NOV-2000; 2000MO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 PI Baker KP, Betesini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-408281/43.
 XX P-PSDB: AAN12252.
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 3; Fig 161; 813pp: English.
 CC AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIa. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 1376 BP; 311 A; 384 C; 352 G; 329 T; 0 other;
 Query Match 18.1%; Score 1117.8; DB 22; Length 1376;
 Best Local Similarity 99.8%; Pred. No. 1.1e-239;
 Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 190 CCAGGCGGAGGCGACGCTGCGCCGGGCTTGGGGCCCTGGGGCCCGCGCTCCACCCTGCG 249
 |||||||
 Db 1 CCAAGGCGGAGGCGACGCTGCGCCGGGCTTGGGGCCCTGGGGCCCGCGCTCCACCCTGCG 60
 QY 250 TTTTCCCAACCGAGGCGGAGGCGTCCCGAGTCAATGCGCGGCTTGAATCTGCGGGTCTCT 309
 |||||||
 Db 61 TTTTCCCAACCGAGGCGGAGGCGTCCCGAGTCAATGCGCGGCTTGAATCTGCGGGTCTCT 120
 QY 310 ATCGACCTGTAGGGGTTTCTGCTGCTGGGTGCGGGCGGCTTGGCGCGGCGGAGAACT 369
 |||||||
 Db 121 ATCGACCTGTAGGGGTTTCTGCTGCTGGGTGCGGGCGGCTTGGCGCGGCGGAGAACT 180
 QY 370 TTTGAGATTGCTGTCCAGAGAGAAACCAATATACAGTTCTCTAAAGCTGGGGACCCCG 429
 |||||||
 Db 181 TTTGAGATTGCTGTCCAGAGAGAAACCAATATACAGTTCTCTAAAGCTGGGGACCCCG 240
 QY 430 ACTCTGCTGGCAAAACCTGTTACATCGTCAATTTCTAAAGACATTAACATGTTGTCC 489
 |||||||
 Db 241 ACTCTGCTGGCAAAACCTGTTACATCGTCAATTTCTAAAGACATTAACATGTTGTCC 300
 QY 490 ATCAAGTCTGGAGAAAGATAGTCTTACCTTTAGCTGCCAGAGTCTGAGAAATCACTT 549
 |||||||
 Db 301 ATCAAGTCTGGAGAAAGATAGTCTTACCTTTAGCTGCCAGAGTCTGAGAAATCACTT 360
 QY 550 GTCATGAGATCCAGAAATATATGACTGATGTCAGGGCCCATGTCCTTTGGGAGGTT 609
 |||||||
 Db 361 GTCATGAGATCCAGAAATATATGACTGATGTCAGGGCCCATGTCCTTTGGGAGGTT 420
 QY 610 CAGCTTCAGCCCTCGACATCGTGTGCTTACCCTACCGTCAAGAACTTTCATCTGGAGTGC 669
 |||||||
 Db 421 CAGCTTCAGCCCTCGACATCGTGTGCTTACCCTACCGTCAAGAACTTTCATCTGGAGTGC 480
 QY 670 AAAGCTCATTAAGACATCGGTTTAAAGCTGACATTTTCCATCCCTGCTGAGGCGATG 729
 |||||||
 Db 481 AAAGCTCATTAAGACATCGGTTTAAAGCTGACATTTTCCATCCCTGCTGAGGCGATG 540
 QY 730 GGTCCGGGTGAGAGTCCGCGACAGAGTCACTCATTCATCGAGCGGCGGAATCGATGCC 789
 |||||||
 Db 541 GGTCCGGGTGAGAGTCCGCGACAGAGTCACTCATTCATCGAGCGGCGGAATCGATGCC 600
 QY 790 ACCGTGTCAGATCGGAACCTTCTGACGAATGAGCACTGTGTCCCGGATCAAGATGCA 849
 |||||||
 Db 601 ACCGTGTCAGATCGGAACCTTCTGACGAATGAGCACTGTGTCCCGGATCAAGATGCA 660
 QY 850 GAAGAGTGAATATGCGCTTACACCTCCATGCTTCCACCCAGAAATGTCTCCGCTTC 909
 |||||||
 Db 661 GAAGAGTGAATATGCGCTTACACCTCCATGCTTCCACCCAGAAATGTCTCCGCTTC 720
 QY 910 AGCATTCGAACCGCTCATCTATAAAGCTGTGATCATGCAAGTGTGTTGAGGGT 969
 |||||||
 Db 721 AGCATTCGAACCGCTCATCTATAAAGCTGTGATCATGCAAGTGTGTTGAGGGT 780
 QY 970 GAAGGCTCAACAACCTGTATGTCTGCCAATACCAAGAGCTTCCCTGAGATGAGCTC 1029
 |||||||
 Db 781 GAAGGCTCAACAACCTGTATGTCTGCCAATACCAAGAGCTTCCCTGAGATGAGCTC 840
 QY 1030 ATGACGTGACGATTTGTCTCTGTGCAACCTGCGGGCAGCGCTCTCTCAACTTC 1089
 |||||||
 Db 841 ATGACGTGACGATTTGTCTCTGTGCAACCTGCGGGCAGCGCTCTCTCAACTTC 900
 QY 1090 AACCTTCCACCTGTAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149
 |||||||
 Db 901 AACCTTCCACCTGTAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 QY 1150 ACCAACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1209
 |||||||
 Db 961 ACCAACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 QY 1210 AACCTCTCTGCAAGGCTGTGACCAAGATGCCCAAGATGCCCAAGATGCCCTCGGCTGCA 1269
 |||||||
 Db 1021 AACCTCTCTGCAAGGCTGTGACCAAGATGCCCAAGATGCCCAAGATGCCCTCGGCTGCA 1080
 QY 1270 TTCCAAGTTTGTGTCACATCCAAATGAAGCAATRA 1310

Db 1081 TTCCAGTTTGTGCACATCCACAAATGAAAGCAGTGA 1121

RESULT 7

ID ABV30293 standard; cDNA: 1031 BP.

XX ABV30293;

DE 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 30284.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KM pharmacogenomic marker; gene; ss.

XX Homo sapiens.

PN W0200160860-A2.

XX 23-AUG-2001.

PF 20-FEB-2001; 2001MO-US05171.

XX 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-235281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1; Page 6580; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 1031 BP; 272 A; 255 C; 212 G; 263 T; 29 other;

XX Query Match 11.7%; Score 718; DB 23; Length 1031;

XX Best Local Similarity 96.5%; Pred. No. 2.1e-150;

XX Matches 745; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

QY 4499 CTTTCTTTTCTTTTAAATGACAGAGATCTATCTGTGCTTACGCTGAGATGC 4558

DB 30 CTTTCTTTTCTTTTAAATGACAGAGATCTATCTGTGCTTACGCTGAGATGC 89

QY 4559 AGTGGCCATCTCGGCTGACTGCAACCTCTGCTCGGCTCAACCATCTCCACAC 4618

DB 90 AGTGGCCATCTCGGCTGACTGCAACCTCTGCTCGGCTCAACCATCTCCACAC 149

QY 4619 TCACCTCCCAAAATAGCTGGATCATCTGCGACAAACCACATGCCAGCTAATTTGTAT 4678

DB 150 TCAGCTCCCAAAATAGCTGGATCATCTGCGACAAACCACATGCCAGCTAATTTGTAT 209

QY 4679 TTTTGTAGAGACAGGCTTTCACCATCTTGGCCAGGCTGCTCACTCTGGGCTCAA 4738

DB 210 TTTTGTAGAGACAGGCTTTCACCATCTTGGCCAGGCTGCTCACTCTGGGCTCAA 269

QY 4739 GCATCTCCCTGCTGGCTGCCCAAGTGCTGGATTTACAGATGTAGGCCACCGCATCC 4798

DB 270 GCATCTCCCTGCTGGCTGCCCAAGTGCTGGATTTACAGATGTAGGCCACCGCATCC 329

QY 4799 AGCCCAACACCTCATTTATATACCAATTAATCTGCCCCATTAATCTGGAATTTGCTCTC 4858

DB 330 AGCCCAACACCTCATTTATATACCAATTAATCTGCCCCATTAATCTGGAATTTGCTCTC 389

QY 4859 ACCCTGCTCTGATCTGGAAGAGAGAGGATTTATGTATAGCTGTGACACAGTCCCAAG 4918

DB 390 ACCCTGCTCTGATCTGGAAGAGAGGATTTATGTATAGCTGTGACACAGTCCCAAG 449

QY 4919 TTCAATATTTCTGGCGCAAAACTTCTCAAAAATTAATATGATCTTCAATTTGTCAT 4978

DB 450 TTCAATATTTCTGGCGCAAAACTTCTCAAAAATTAATATGATCTTCAATTTGTCAT 509

QY 4979 GAATTCACCTTGGAAATGACACCGCTCACTTGTTCACATGACATTAATGAAGAAATTT 5038

DB 510 GAATTCACCTTGGAAATGACACCGCTCACTTGTTCACATGACATTAATGAAGAAATTT 569

QY 5039 TATAGTCTCTTAATGAGCGGTGATCTGCAAGACCTTGAACACTTTCAGAGATAGAT 5098

DB 570 TATAGTCTCTTAATGAGCGGTGATCTGCAAGACCTTGAACACTTTCAGAGATAGAT 629

QY 5099 ATTTAGTCAATGACCTTGGCGTTGCTGATGACACTTCCCTTGAAGCTGGTCTCT 5158

DB 630 ATTTAGTCAATGACCTTGGCGTTGCTGATGACACTTCCCTTGAAGCTGGTCTCT 689

QY 5159 GCCCAGTACCTTGGCTTGTGAGCCGAGATGCTGACCTGCAATTAAGGCCAAGAG 5218

DB 690 GCCCAGTACCTTGGCTTGTGAGCCGAGATGCTGACCTGCTT--AGGCCAAGAGAG 747

QY 5219 GCGTGGGCTTCTTCCCTCAGTGAAGCCCTTATTTGATTAATGATGCTG 5270

DB 748 GCGTGGGCTTCTTCCCTCAGTGAAGCCCTTATTTGATTAATGATGCTG 799

XX RESULT 8

XX AAH08379 standard; cDNA: 659 BP.

XX AAH08379;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:5214.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1: SEQ ID 5214; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH0166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 659 BP; 150 A; 187 C; 168 G; 151 T; 3 other:
SQ
Query Match 10.2%; Score 630.8; DB 22; Length 659;
Best Local Similarity 98.9%; Pred. No. 5.4e-131;
Matches 654; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

OY 1049 TTCCTGACACCTGGCGGCGAGCTGCTTCTCACTTCAACCTCTCCAACTGTGAGA 1108
|||||
DB 480 TTCCTGACACCTGGCGGCGAGCTGCTTCTCACTTCAACCTCTCCAACTGTGAGA 539
OY 1109 GGAAGAGGAGCGGGTGTGAATACATCCGGGCTCCACACCAACCCGAGGTGTGA 1168
|||||
DB 540 GGAAGAGGAGCGGGTGTGAATACATCCGGGCTCCACACCAACCCGAGGTGTGA 599
OY 1169 AGCTGGAGGACAAAGCAGCTGGGACATGGCGGGGAACTTCAACTCTCTGCAAGGCT 1228
|||||
DB 600 A-CTGGAGGACAAAGCAGCTGGGACATGGCGGGGAACTTCAACTCTCTGCAAGGCT 658
OY 1229 G 1229
DB 659 G 659
RESULT 9
AAA78075
ID AAA78075 standard; cDNA; 544 BP.
XX
XX AAA78075;
AC
XX
XX 14-NOV-2000 (first entry)
DT
XX
XX cDNA encoding human colon tumour polypeptide. SEQ ID NO:362.
DE
XX
XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW Immunotherapy; diagnosis; progression; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200037643-A2.
PN
XX
XX 29-JUN-2000.
PD
XX
XX 23-DEC-1999; 99WO-US30909.
PF
XX
XX 23-DEC-1998; 98US-0221298.
PR
XX 02-JUL-1999; 99US-0347496.
PR 22-SEP-1999; 99US-0401064.
PR 19-NOV-1999; 99US-0444242.
PR 02-DEC-1999; 99US-0454150.
XX
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yqñu J;
PI
XX
XX WPI: 2000-442671/38.
DR
XX
XX New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -
PT
XX
XX Claim 1: Page 196; 229pp; English.
PS
XX
XX Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours.
CC The invention also specifically discloses 8 human colon tumour proteins
CC (AAH1897-B11904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs), preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specifically for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer

CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level
 CC of a tumour protein of the invention, and therefore to determine whether
 CC cancer cells are present. Such diagnostic methods may also be used to
 CC monitor the progression of a cancer by repeating the processes at time
 CC intervals, and comparing the current result to previous results. The
 CC present sequence represents a cDNA encoding a human colon tumour
 CC polypeptide.

XX Sequence 544 BP; 138 A; 123 C; 113 G; 170 T; 0 other;

Query Match 8.8%; Score 542.4; DB 21; Length 544;
 Best Local Similarity 99.8%; Pred. No. 2.8e-111;

Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3068 CCTGAGTCACCATGATAGGTTGTCAGACAGCCCTGGATTTCAGAGTTTAAACAGAGCT 3127
 DB 1 CCTGAGTCACCATGATAGGTTGTCAGACAGCCCTGGATTTCAGAGTTTAAACAGAGCT 60

QY 3128 TGGCCCTTTCAGACAAACAGTTCATTCAGAGAGCCCTGAGTCCCTACTCTCAC 3187
 DB 61 TGGCCCTTTCAGACAAACAGTTCATTCAGAGAGCCCTGAGTCCCTACTCTCAC 120

QY 3188 TGGGGTCCCGAGATGAAAAGCAATGCGCTTTTATATTATTATTGTTGTTGCT 3247
 DB -121 TGGGGTCCCGAGATGAAAAGCAATGCGCTTTTATATTATTATTGTTGTTGCT 180

QY 3248 GTGTTATTAAAGATCAAAATGTTAAACACCTAGCTCTTTTACCTGACTTAAATAA 3307
 DB 181 GTGTTATTAAAGATCAAAATGTTAAACACCTAGCTCTTTTACCTGACTTAAATAA 240

QY 3308 CTCTACTAACTGTTTGGATGCTGGGTTGTGACTTCTACTGACGCCCTGATTAACGTG 3367
 DB 241 CTCTACTAACTGTTTGGATGCTGGGTTGTGACTTCTACTGACGCCCTGATTAACGTG 300

QY 3368 TGCCCTGCCCCAGGTGGTGGGAATTAATTAACAATCTGTCCAAACAGAAAAGATGCTG 3427
 DB 301 TGCCCTGCCCCAGGTGGTGGGAATTAATTAACAATCTGTCCAAACAGAAAAGATGCTG 360

QY 3428 TGTGTGACGACATGACACATATCTGTTGATTAAGAGACTTCCGATTTCTAGCTG 3487
 DB 361 TGTGTGACGACATGACACATATCTGTTGATTAAGAGACTTCCGATTTCTAGCTG 420

QY 3488 GTTCGTGTTATCCCATTTGGAATTCATCTGAATCCATTTCTATATAGTCTAGCA 3547
 DB 421 GTTCGTGTTATCCCATTTGGAATTCATCTGAATCCATTTCTATATAGTCTAGCA 480

QY 3548 ATTAAGAGAAATTCCTCAAGTTTCCATGTGCGGTTCTCTAGCTGACGCAATCTTGAC 3607
 DB 481 ATTAAGAGAAATTCCTCAAGTTTCCATGTGCGGTTCTCTAGCTGACGCAATCTTGAC 540

QY 3608 ATTT 3611
 DB 541 ATTT 544

RESULT 10
 AA128813
 ID AA128813 standard; cDNA; 544 BP.

XX AA128813;
 XX 12-OCT-2001 (first entry)

DE Colon tumour related determined cDNA sequence for clone 25928.
 XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW Immunogenic; gene therapy; vaccine; colonic cancer; ss.

XX Homo sapiens.
 XX W0200149716-A2.
 PN
 XX

PD 12-JUL-2001.

XX 29-DEC-2000; 2000NOV-0535596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 15-FEB-2000; 2000US-0504629.

XX 06-MAR-2000; 2000US-0519444.

XX 19-MAY-2000; 2000US-0575251.

XX 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes M, Secret H, Benson DR, Meagher M, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 PI WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
 PS prevention, diagnosis and treatment of colonic cancer -
 Claim 2; Page 233; 472pp; English.

CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patient's own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA124494 to AA124523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 544 BP; 138 A; 123 C; 113 G; 170 T; 0 other;

Query Match 8.8%; Score 542.4; DB 22; Length 544;
 Best Local Similarity 99.8%; Pred. No. 2.8e-111;
 Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3068 CCTGAGTCACCATGATAGGTTGTCAGACAGCCCTGGATTTCAGAGTTTAAACAGAGCT 3127
 DB 1 CCTGAGTCACCATGATAGGTTGTCAGACAGCCCTGGATTTCAGAGTTTAAACAGAGCT 60

QY 3128 TGGCCCTTTCAGACAAACAGTTCATTCAGAGAGCCCTGAGTCCCTACTCTCAC 3187
 DB 61 TGGCCCTTTCAGACAAACAGTTCATTCAGAGAGCCCTGAGTCCCTACTCTCAC 120

QY 3188 TGGGGTCCCGAGATGAAAAGCAATGCTCTTTTATTATTATTATTATTGTTGTTGCT 3247
 DB 121 TGGGGTCCCGAGATGAAAAGCAATGCTCTTTTATTATTATTATTATTGTTGTTGCT 180

QY 3248 GTGTTATTAAAGATCAAAATGTTAAACACCTAGCTCTTTTACCTGACTTAAATAA 3307
 DB 181 GTGTTATTAAAGATCAAAATGTTAAACACCTAGCTCTTTTACCTGACTTAAATAA 240

QY 3308 CTCTACTAACTGTTTGGATGCTGGGTTGTGACTTCTACTGACGCCCTGATTAACGTG 3367
 DB 241 CTCTACTAACTGTTTGGATGCTGGGTTGTGACTTCTACTGACGCCCTGATTAACGTG 300


```
XX
DR WPI, 2001-611627/70.
XX
PT New colon tumour proteins and related nucleic acid, useful for
PT treatment, prevention, diagnosis and monitoring of cancer -
XX
PS Claim 4; Page 162; 299pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC encoding for at least an immunogenic portion of human colon tumour
CC proteins. The sequences of the invention are useful in pharmaceutical
CC compositions and vaccines for the prevention and treatment of cancers
CC such as colon cancer. They are also useful for the diagnosis and
CC monitoring of such cancers. Antibodies to the colon tumour proteins
CC and antigen presenting cells that express polynucleotides encoding
CC colon tumour proteins can be used to inhibit the development of
CC cancers. T-cells that react specifically with colon tumour proteins
CC are useful for removing tumour cells from samples (e.g. blood) and
CC for cancer treatment. The polynucleotide sequences are also useful in
CC gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the
CC invention that encode for portions of human colon tumour proteins.
XX
SQ Sequence 636 BP; 153 A; 150 C; 135 G; 186 T; 12 other;

Query Match      8.5%; Score 522.6; DB 23; Length 636;
Best Local Similarity 98.3%; Pred. No. 7.8e-107;
Matches 536; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 3068 CCTAGTACCTAGTACGAGGTTGACAGCCCTGGATTGAGTGTAAACAGAGGCT 3127
DB 1 CCTAGTACCTAGTACGAGGTTGACAGCCCTGGATTGAGTGTAAACAGAGGCT 60
QY 3128 TGCCCTCTGAGAGCAACAGTTCATTCAGAGAGCCTGAGGCTCCCTACTCTCAC 3187
DB 61 TGCCCTCTGAGAGCAACAGTTCATTCAGAGAGCCTGAGGCTCCCTACTCTCAC 120
QY 3188 TGGGGTCCCAAGATGAAAACGACATGTCCTTTATATATTTATTTGTTGCTCT 3247
DB 121 TGGGGTCCCAAGATGAAAACGACATGTCCTTTATATATTTATTTGTTGCTCT 180
QY 3248 GTGTATTTAAGATCAAAATGTATTAACCACTGCTTTTCCGACTTATATATA 3307
DB 181 GTGTATTTAAGATCAAAATGTATTAACCACTGCTTTTCCGACTTATATATA 240
QY 3308 CTGATTAATGCTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3367
DB 241 CTGATTAATGCTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 3368 TGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3427
DB 301 TGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 3428 TGTGTGAGAGCATGACATATCTGCTTGTATGAGAGCTTCTGATTCCTAGGTCG 3487
DB 361 TGTGTGAGAGCATGACATATCTGCTTGTATGAGAGCTTCTGATTCCTAGGTCG 420
QY 3488 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3547
DB 421 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 3548 ATAAAGAGAAATTTCTCAAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTG 3607
DB 481 ATAAAGAGAAATTTCTCAAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTG 539
QY 3608 ATTTA 3612
DB 540 ATTTA 544

RESULT 13
AAH13321/c
ID AAH13321 standard; cDNA; 561 BP.
XX
```

```
AC AAH13321;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (3'-primer) SEQ ID NO:10156.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 3; SEQ ID 10156; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination
XX of the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 561 BP; 161 A; 136 C; 120 G; 140 T; 4 other;

Query Match      8.3%; Score 509.4; DB 22; Length 561;
Best Local Similarity 96.1%; Pred. No. 6.6e-104;
Matches 541; Conservative 0; Mismatches 20; Indels 2; Gaps 2;

QY 5581 TCAGGTCACATGGAACAGGAGTGCATTTGTTGAGGCTTACTGATCCAGCCACT 5640
DB 561 TCAGGTCACATGGAACAGGAGTGCATTTGTTGAGGCTTACTGATCCAGCCACT 503
QY 5641 GCTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5700
DB 502 GCTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
```

QY	5701	AATGCTGGGCAATTTGGCAGCAAGAAAGGAAAGTCTCAAGGAAAGGTCAACTGGGACAAAGCA	5760
Db	443	AATGCTGGGCAATTTGGCAGCAAGAAAGGAAAGTCTCAAGGAAAGGTCAACTGGGACAAAGCN	384
QY	5761	CAAGCCCAACCGACATGGCCCTTGGTAAAGGTTAGCAGACTGTGTGTGGATCTGCAGT	5820
Db	383	CAAGCCCAACCGACATGGCCCTTGGTAAAGGTTAGCAGACTGTGTGTGGATCTGCAGT	324
QY	5821	GCTTCACTGGAAATTAATTATTCATTGCAGATACCTTTTAGTGGGCAATTTATTCATTTTC	5880
Db	323	GCTTCACTGGAAATTAATTATTCATTGCAGATACCTTTTAGTGGGCAATTTATTCATTTTC	264
QY	5881	CTGGCTTTAAATTAACAAATGTACCAAAAAACAAGATACAGCTGTTTAAAGCTTCGG	5940
Db	263	CTGGCTTTAAATTAACAAATGTACCAAAAAACAAGATACAGCTGTTTAAAGCTTCGG	204
QY	5941	CTACTTGTCCCCCTGGTTCAGTAGAGGCCCGGTTTCCAGTTGTGACTGTGACAGGCTC	6000
Db	203	CTACTTGTCCCCCTGGTTCAGTAGAGGCCCGGTTTCCAGTTGTGACTGTGACAGGCTC	144
QY	6001	AGCATGGGCTCAGCAGATGCTGTCTTAATTTGTGGATGATACAGAAACCAAGGCTTTGGG	6060
Db	143	AGCATGGGCTCAGCAGATGCTGTCTTAATTTGTGGATGATACAGAAACCAAGGCTTTGGG	84
QY	6061	ATCAAGAGTCTTTCCTCTCATTTTGAGCCGTCGACAGCTGTGAAGCAGATGTTTTGTCC	6120
Db	83	ATCAAGAGTCTTTCCTCTCATTTTGAGCCGTCGACAGCTGTGAAGCAGATGTTTTGTCC	24
QY	6121	GGAATATATAATTAATAGCTTGG 6143	
Db	23	GGAATATATAATTAATAGCTTGG 1	
RESULT 14			
ID	AAZ80286/C		
AAZ80286	standard; cDNA; 636 BP.		
XX	AAZ80286;		
XX	07-APR-2000 (first entry)		
DE	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:370.		
KW	Human; gene expression product; diagnosis; tumour; colon cancer;		
KM	colorectal adenocarcinoma; cell line SW480; cell proliferation;		
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;		
XX	hyperplasia; ds.		
OS	Homo sapiens.		
PN	WC0964576-A2.		
XX	16-DEC-1999.		
XX	09-JUN-1999; 99WO-IB01062.		
PR	10-JUN-1998; 98US-0088801.		
PA	(FARB) BAYER CORP.		
PI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;		
PI	Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;		
PI	Schlegel R;		
DR	WPI: 2000-087220/07.		
XX			
XX	Novel nucleic acids, used to develop products for the diagnosis and		
PT	treatment of disorders involving unwanted cell proliferation,		
PT	particularly cancers, especially colon cancer		
XX			
PS	Claim 15; Page 280; 469pp: English.		
CC	AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from		

CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cell
CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.

XX Sequence 636 BP; 155 A; 144 C; 155 G; 160 T; 22 other;

XX

Query Match	7.98;	Score 489.2;	DB 21;	Length 636;
Best Local Similarity	90.38;	Pred. No. 2.3e-99;		
Matches 577;	Conservative	0;	Mismatches 54;	Indels 8;
				Gaps 6;

QY	5270	GGAGCCCTTACGCCCTCCATTCCTGCACAAATTCGCCCAACCTCCAGCCCTCCCAAGCAGACT	532	39
Db	636	GGAAAGCCNANNCTCCCACTTTTAAANATTCCAACCTT--CAAGCCCTTCCAAACAGSANT	579	
QY	5330	AGGTGCCCTGCATTCACACCCCAAGGTGGGATTTGGCTTCTCTT--AGGCTGGCTACTTG--TC	538	
Db	578	A-GAGCCCTNGANTCCNCCCAAGGNGGATTTGGCTTCCCTTAAGSNGNGTACTTATGTC	520	
QY	5387	ACCATCACCGACATCACTGTGGCTTCGACAGGACACACGTGGCATTTTCTTC--AACTG	545	
Db	519	ACNATCAGGAAATNNCTGGTGGCTTCANAGNACNCCAAAGGGGCCATTTTCTTCAAAATA	460	
QY	5446	AGGAGCTCAAAAACCTCTGGACAATGCTGGCTCC--TGAACACAGATATTTCCTGAGACTGT	550	
Db	459	AGGGCTCAAAACCTCTGGACAATGCTGGCTCCTTAAGNCCAGATTTTCTCGAAGCTGG	400	
QY	5505	GCCTCAGTGAAGGGGCCAGCCTTGAGAACCTTGCTCTTTTCTTTAAAGCCAGGCC	556	
Db	399	CCTCAAGTAAAGGGGCCCACTTGAGGAACCCCTGGCTCTTTCTTTTAAAGCCAGGCC	340	
QY	5565	ACTTACATPAAAACTTTACAGGGTCACTGGAAACAGTGAAGTGCATTTGTTGAAGCTTAC	562	
Db	339	ACTTACGTAATAACATTTTCAGGGTCACTGGAAACAGTGAAGTGCATTTGTTGAAGCTTAC	280	
QY	5625	TGCATGCCAGCCCACTGCTCATCCAGGTGGTCTGCCATGCTACGAGGAAGGCCAGCGCA	568	
Db	279	TGCATGCCAGCCCACTGCTCATCCAGGTGGTCTGCCATGCTACGAGGAAGGCCAGCGCA	220	
QY	5685	TGCAGGACTGCTCTTAATGCTGTGTCATGTGCACAGAAAGGAAAGTCTCAAGAAAG	574	
Db	219	TGCAGGACTGCTCTTAATGCTGTGTCATGTGCACAGAAAGGAAAGTCTCAAGAAAG	160	
QY	5745	TCACTGGGACAAGCACAAAGCCCAAGCATGGCTTGGTAAAGATTAGCAGACTGGTG	580	
Db	159	TCACTGGGACAAGCACAAAGCCCAAGCATGGCTTGGTAAAGATTAGCAGACTGGTG	100	
QY	5805	TGTGTGATCTGCAGTGTCTCACTGGAATATTTATTCATTTGCAGATACATTTTAAAGTG	586	
Db	99	TGTGTGATCTGCAGTGTCTCACTGGAATATTTATTCATTTGCAGATACATTTTAAAGTG	40	
QY	5865	GCATTTATTCATTTCTGTGCTTAAATTAACAATATG	590	
Db	39	GCATTTATTCATTTCTGTGCTTAAATTAACAATATG	1	
RESULT 15				
AAS64483				
ID AAS64483 standard; cDNA; 487 BP.				
XX AAS64483;				
XX 13-FEB-2002 (first entry)				
XX DNA encoding novel human diagnostic protein #287.				
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;				
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.				

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 09:32:51 : Search time 63 Seconds
(without alignments)
1768.215 Million cell updates/sec

Title: US-09-899-569a-4

Perfect score: 4394
Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKQTDIPLTNTQPMERAE 836

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4394	100.0	836	23	AA049641 Human tumour-assoc
2	3634.5	82.7	709	21	AA049645 Human secreted pro
3	3467	78.9	749	23	AA049640 Human tumour-assoc
4	3411	77.6	649	22	AA055558 Human protein sequ
5	2155	49.0	443	21	AA049592 Human secreted pro
6	1779	40.5	343	22	AA012252 Human PRO5773 poly
7	827	18.8	159	21	AA049593 Human secreted pro
8	775	17.6	143	22	ABG00296 Novel human diagno
9	616	14.0	121	22	AA023969 Human EST encoded
10	523	11.9	167	22	AA059483 Human purified sec

11	390	8.9	116	22	AA090342 Human immune/haema
12	140.5	3.2	3623	20	AA027020 Rat cubilin proteol
13	126	2.9	595	22	AA070070 Human secreted pro
14	126	2.9	595	23	ABG65509 Human albumin fusi
15	119	2.7	3095	23	AA020788 Rat C3b/C4b comple
16	118.5	2.7	1406	21	AA042916 Human ORFX ORF2680
17	118	2.7	1733	23	AB006023 Human cubilin prot
18	118	2.7	2118	22	ABG23416 Novel human diagno
19	118	2.7	3631	22	ABG28277 Novel human diagno
20	117.5	2.7	774	22	ABG64637 Drosophila melanog
21	117.5	2.7	774	22	AA038959 Drosophila G-prote
22	117.5	2.7	1464	21	AB071111 Drosophila melanog
23	117.5	2.7	1719	17	AA092100 Human RIZ allele D
24	117	2.7	1045	23	AB091371 Herbicidally activ
25	117	2.7	2476	20	AA067738 Pig p105 zona pell
26	116.5	2.7	4215	22	ABG59065 Drosophila melanog
27	116	2.6	1890	23	AB060622 Human cubilin prot
28	114	2.6	477	23	AA023674 SIG protein. Unid
29	113.5	2.6	838	22	AB070548 Drosophila melanog
30	113	2.6	3329	21	AA056029 Human BRCA2 tumour
31	112.5	2.6	433	17	AA098454 Human oligodendrocyte-my
32	112.5	2.6	433	21	AA057091 Human bone marrow
33	112.5	2.6	581	22	AB012455 Human polyepitide
34	112.5	2.6	1770	22	AA039102 Human semaphorin G
35	112.5	2.6	1805	22	ABG04010 Human human diagno
36	112	2.5	736	23	AB057361 Mouse ischaemic co
37	112	2.5	1848	22	AA040888 Human polyepitide
38	111.5	2.5	728	22	AA085060 Human serine prote
39	111.5	2.5	1393	22	AA078991 Human protein SEQ
40	110.5	2.5	771	23	AA068311 Human semaphorin G
41	110.5	2.5	870	23	AA068303 Human semaphorin G
42	110.5	2.5	942	23	AB091468 Herbicidally activ
43	110.5	2.5	1694	21	AA046461 Rat laminin 5 poly
44	110.5	2.5	1719	21	AA012029 Human Rb-interacti
45	110.5	2.5	1719	21	AA012113 Human Rb-interacti

ALIGNMENTS

RESULT 1
AA049641 standard: Protein; 836 AA.
AA049641:
17-MAY-2002 (first entry)
Human tumour-associated antigen B345 protein SEQ ID NO 4.
Tumour-associated antigen; human; B345; cytostatic; cell communication;
cell interaction; signal transduction; metastasis; cancer; colon;
immunotherapy; carcinoma; lung; diagnosis.
Homo sapiens.
WO200204508-A1.
17-JAN-2002.
05-JUL-2001; 2001WO-EP07705.
07-JUL-2000; 2000DE-1033080.
PR 19-APR-2001; 2001DE-1019294.
(BOEH) BOEHRINGER INGELHEIM INT GMBH.
Schweifer N, Scherl-Mostageer W, Sommergruber W, Adseher R;
WPI; 2002-171704/22.
N-PSDB; ABA959507.
New tumour-associated antigen B345, useful for diagnosis and

PT immunotherapy of tumors, also related nucleic acid and antibodies -
 XX
 XX Claim 1; Page 85-88; 102pp; German.
 CC This invention describes a novel tumour-associated antigen, designated
 CC B345 which has cytostatic activity. B345 is involved in communication,
 CC interaction and/or signal transduction with extracellular components and
 CC ligands, especially in the metastatic potential of cancers, particularly
 CC of the colon. B345 or its immunogenic fragments, also the DNA that
 CC encodes it, are useful for immunotherapy of cancer, particularly
 CC carcinoma of lung or colon. Antibodies raised against B345 are useful for
 CC treatment and diagnosis of cancers that are associated with B345
 CC expression, including their use for targeted delivery of cytotoxic or
 CC radioactive agents. Probes derived from B345 can be used to detect
 CC tumour-specific mutations in the B345 sequence, and can be used to screen
 CC for B345 specific modulators. This sequence represents a human B345
 CC tumour-associated antigen described in the invention.
 CC
 XX
 SQ Sequence 836 AA:
 Query Match 100.0%; Score 4394; DB 23; Length 836;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGLNCVSTALLGVLGAAALPGCAEALPRESNTVLKIGTPPLAKPCYIV 60
 DB 1 MAGLNCVSTALLGVLGAAALPGCAEALPRESNTVLKIGTPPLAKPCYIV 60
 QY 61 SKRHITMLSTKSGRIYFTSCSPENHFYIEIOKNIDCSGPCPEEVLQSTSLPT 120
 DB 61 SKRHITMLSTKSGRIYFTSCSPENHFYIEIOKNIDCSGPCPEEVLQSTSLPT 120
 QY 121 LNRTFMDVAKHSIGLELOFSIPRLQIGPGSCPDGVTSHISGRIDATVVRIGTECSN 180
 DB 121 LNRTFMDVAKHSIGLELOFSIPRLQIGPGSCPDGVTSHISGRIDATVVRIGTECSN 180
 QY 181 GTVSRKIMQGVMAHLPMFHPRANVSGESIANBSIKRLCTIESVREGSATLMSANT 240
 DB 181 GTVSRKIMQGVMAHLPMFHPRANVSGESIANBSIKRLCTIESVREGSATLMSANT 240
 QY 241 PEGPPEDELTMQFVPAHLRASVFLNMLSCERKEEVEYIIPSTNPEFTEKDK 300
 DB 241 PEGPPEDELTMQFVPAHLRASVFLNMLSCERKEEVEYIIPSTNPEFTEKDK 300
 QY 301 QPQNMAGNENLSTQGCDDQASPGILRLQFVYVQHPQNSNKIYVVDLSNERAMSLTIE 360
 DB 301 QPQNMAGNENLSTQGCDDQASPGILRLQFVYVQHPQNSNKIYVVDLSNERAMSLTIE 360
 QY 361 PRVYKQSRKVPCCFVCLSEKRTSSNLTLSGSKHKISFLCDDLRLMNVKRTISCTDH 420
 DB 361 PRVYKQSRKVPCCFVCLSEKRTSSNLTLSGSKHKISFLCDDLRLMNVKRTISCTDH 420
 QY 421 RYQQRKSYSLQVPSDLHLRVELHDFSWKLLVPRKDRSLVLPAOKQOHTHKPCQTSF 480
 DB 421 RYQQRKSYSLQVPSDLHLRVELHDFSWKLLVPRKDRSLVLPAOKQOHTHKPCQTSF 480
 QY 481 SYLVASAIPIPSQDLVFCGSGSIIKOIYQKNTSVTLTRFAPSFQOASNOGLTVSITPY 540
 DB 481 SYLVASAIPIPSQDLVFCGSGSIIKOIYQKNTSVTLTRFAPSFQOASNOGLTVSITPY 540
 QY 541 FKEEGVFTVPTDKSKVYLLTPMWDKGLPSLTVSNWISVPRQVACLFFKERSGVYCO 600
 DB 541 FKEEGVFTVPTDKSKVYLLTPMWDKGLPSLTVSNWISVPRQVACLFFKERSGVYCO 600
 QY 601 TGAFPMIOQORTRAEITFLDEDVLPKPSFHHHSFVWNISNCSPTSGKDLDFSTLT 660
 DB 601 TGAFPMIOQORTRAEITFLDEDVLPKPSFHHHSFVWNISNCSPTSGKDLDFSTLT 660
 QY 661 PRVVDLTVLIIAIVGGVLLLSALGLITICVKKKKKTKNGPVGITNGINTEMPQPK 720
 DB 661 PRVVDLTVLIIAIVGGVLLLSALGLITICVKKKKKTKNGPVGITNGINTEMPQPK 720
 QY 721 KFGQGRKNDNSHVAVIEDTMVYGHLLQDSSGSFLQPEVDIYRPFQGTGVCPPSPPTIC 780

DB 721 KFGQGRKNDNSHVAVIEDTMVYGHLLQDSSGSFLQPEVDIYRPFQGTGVCPPSPPTIC 780
 QY 781 SRAPTAKLATEPPRPSPSESEPTFTSHPNNDVSKKTDIPLNTQPMERAE 836
 DB 781 SRAPTAKLATEPPRPSPSESEPTFTSHPNNDVSKKTDIPLNTQPMERAE 836
 RESULT 2
 AA91456
 ID AA91456 standard; Protein; 709 AA.
 XX
 AC AA91456;
 XX
 DT 29-JUN-2000 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 6 SHQ ID NO:129.
 XX
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antisthma;
 KW immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
 KW antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 KW antisthma; antiparasitic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC
 XX
 PI Komatsu, I. S. GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
 XX
 DR WPI: 2000-195282/17.
 XX N-PSDB: AAA26351.
 PT New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX
 PS Claim 11; Page 456-459; 634pp; English.
 XX
 CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AA91451 to AA91651. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
 CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 CC antisthma; antiparasitic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

CC cardiovascular disorders, reproductive disorders, gastrointestinal disorders, respiratory disorders and metabolic disorders. The proteins or polynucleotides can also be used as food additives or preservatives. CC The proteins are also useful for identifying their binding partners. CC AA26337 to AA26345 and AA291450 are sequences used in the CC exemplification of the present invention.

Sequence 709 AA:

Query Match 82.7%; Score 3634.5; DB 21; Length 709;
Best Local Similarity 98.5%; Pred. No. 6.1e-308;
Matches 699; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

```

OY 1 MAGLNGVSIALLGVLLGAARLPKGAFAETALPRESNTVILKIKGTPTLLAKPCYIYI 60
DB 1 MAGLNGVSIALLGVLLGAARLPKGAFAETALPRESNTVILKIKGTPTLLAKPCYIYI 60
OY 61 SKRHITMLSIKSGERIVFFSCQSPENHVEIQRNIDCMGSPCPGEVQLQPTSLPT 120
DB 61 SKRHITMLSIKSGERIVFFSCQSPENHVEIQRNIDCMGSPCPGEVQLQPTSLPT 120
OY 121 LMRITMDVKAHKSIGLEQFSIPRLQIQPEBSCPDGYTHSISGRIDATVVRIGTCSN 180
DB 121 LMRITMDVKAHKSIGLEQFSIPRLQIQPEBSCPDGYTHSISGRIDATVVRIGTCSN 180
OY 181 GIVSRITKMQEGVMALHLPWHPRNVSGFSIANRSSIKRLCTIESVEBEGSNTILMSAN 240
DB 181 GIVSRITKMQEGVMALHLPWHPRNVSGFSIANRSSIKRLCTIESVEBEGSNTILMSAN 240
OY 241 PGCFPEDELMTQFVPAHLRASVSFLNPNLSNCKEKEVEYIIGSTNPEVLEDK 300
DB 241 PGCFPEDELMTQFVPAHLRASVSFLNPNLSNCKEKEVEYIIGSTNPEVLEDK 300
OY 301 QPGNAGNPNLSLQCCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTE 360
DB 301 QPGNAGNPNLSLQCCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTE 360
OY 361 PRPVKSRKFPVGCFCLESRTCSNLTLSGSKHISFLCDLRLMNVKETSCTPH 420
DB 361 PRPVKSRKFPVGCFCLESRTCSNLTLSGSKHISFLCDLRLMNVKETSCTPH 420
OY 421 RYCORKSVSLQVPSDILHLPVELHDFSMKLVLPKDRSLVLPKQLOQTHHEKPCNTSF 480
DB 421 RYCORKSVSLQVPSDILHLPVELHDFSMKLVLPKDRSLVLPKQLOQTHHEKPCNTSF 480
OY 481 SYLVASAIRSODLYFGSGFCGSIKOIQVKNISVTLRTFAPSFRQEASRQGLTVSFIY 540
DB 481 SYLVASAIRSODLYFGSGFCGSIKOIQVKNISVTLRTFAPSFRQEASRQGLTVSFIY 540
OY 541 FKEBGFVTVPPTKSKVYLRTPNMDGLPSLTSVSNINISVPRDOVACLFFEKERSVQO 600
DB 541 FKEBGFVTVPPTKSKVYLRTPNMDGLPSLTSVSNINISVPRDOVACLFFEKERSVQO 600
OY 601 TGRAFMIIOEQRTAEIISLDEVDLPKPSFHHNSFWNINISCPISGKQLDLFFSVTLT 660
DB 601 TGRAFMIIOEQRTAEIISLDEVDLPKPSFHHNSFWNINISCPISGKQLDLFFSVTLT 660
OY 661 PRTVDLTVLILAAVGGVLLSALGLITCCVKKKKKTKNGRAVGIYNGN 710
DB 661 PRTVDLTVLILAAVGGVLLSALGLITCCVKKKKKTKNGRAVGIYNGN 710

```

RESULT 3

AA26340 standard; Protein: 749 AA.

AA26340;

17-MAY-2002 (first entry)

Human tumour-associated antigen B345 protein.

Tumour-associated antigen; human; B345; cytosolic; cell communication;

KW cell interaction; signal transduction; metastasis; cancer; colon; immunotherapy; carcinoma; lung; diagnosis.

OS Homo sapiens.

PN WO200204508-A1.

17-JAN-2002.

05-JUL-2001; 2001WO-BP07705.

07-JUL-2000; 2000DE-1033080.

19-APR-2001; 2001DE-1019294.

(BOEH) BOEHRINGER INGELHEIM INT GMBH.

Schweifer N, Scherl-Mostagser M, Sommergruber W, Abseher R;

WPI: 2002-171704/22.

N-PSDB: ABA95506.

New tumour-associated antigen B345, useful for diagnosis and immunotherapy of tumors, also related nucleic acid and antibodies

Example 6; Page 76-78; 102pp; German.

This invention describes a novel tumour-associated antigen, designated B345 which has cytostatic activity. B345 is involved in communication, interaction and/or signal transduction with extracellular components and ligands, especially in the metastatic potential of cancers, particularly of the colon. B345 or its immunogenic fragments, also the DNA that encodes it, are useful for immunotherapy of cancer, particularly carcinoma of lung or colon. Antibodies raised against B345 are useful for treatment and diagnosis of cancers that are associated with B345 expression, including their use for targeted delivery of cytotoxic or radioactive agents. Probes derived from B345 can be used to detect tumour-specific mutations in the B345 sequence, and can be used to screen for B345 specific modulators. This sequence represents a human B345 tumour-associated antigen described in the invention.

Sequence 749 AA:

Query Match 78.9%; Score 3467; DB 23; Length 749;
Best Local Similarity 90.3%; Pred. No. 2.8e-293;
Matches 676; Conservative 9; Mismatches 28; Indels 36; Gaps 6;

```

OY 67 MLISKGERIVFFSCQSPENHVEIQRNIDCMGSPCPGEVQLQPTSLPTLNRPTI 126
DB 1 MLISKGERIVFFSCQSPENHVEIQRNIDCMGSPCPGEVQLQPTSLPTLNRPTI 126
OY 127 WDVKAHKSIGLEQFSIPRLQIQPEBSCPDGYTHSISGRIDATVVRIGTCSNGTVSR 186
DB 127 WDVKAHKSIGLEQFSIPRLQIQPEBSCPDGYTHSISGRIDATVVRIGTCSNGTVSR 186
OY 187 KMOEGVMALHLPWHPRNVSGFSIANRSSIKRLCTIESVEBEGSATILMSANYPGPE 246
DB 187 KMOEGVMALHLPWHPRNVSGFSIANRSSIKRLCTIESVEBEGSATILMSANYPGPE 246
OY 247 DELMTQFVPAHLRASVSFLNPNLSNCKEKEVEYIIGSTNPEVLEDKQCGNNA 306
DB 247 DELMTQFVPAHLRASVSFLNPNLSNCKEKEVEYIIGSTNPEVLEDKQCGNNA 306
OY 306 GNFNLSLQCCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTEPRPVK 366
DB 306 GNFNLSLQCCDDQDQSPGILRLQFVLYOHQPNESKRIYVVDLSNERAMSLTEPRPVK 366
OY 367 SRKFPVGCFCLESRTCSNLTLSGSKHISFLCDLRLMNVKETSCTDHRCQK 426
DB 367 SRKFPVGCFCLESRTCSNLTLSGSKHISFLCDLRLMNVKETSCTDHRCQK 426
OY 427 SYSLQVPSDILHLPVELHDFSMKLVLPKDRSLVLPKQLOQTHHEKPCNTSFSLVS 486
DB 427 SYSLQVPSDILHLPVELHDFSMKLVLPKDRSLVLPKQLOQTHHEKPCNTSFSLVS 486

```


DT 29-JUN-2000 (first entry)
XX Human secreted protein sequence encoded by gene 6 SEQ ID NO:265.
DE
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
KW osteopathic; antiallergic; antibacterial; antidiabetic; antitachy;
KW antipruritic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour.
XX
OS Homo sapiens.
XX
PN WO200006698-A1.
XX
XX 10-FEB-2000.
PD
XX 29-JUL-1999; 99WO-US17130.
XX
XX 30-JUL-1998; 98US-0094657.
PR 05-AUG-1998; 98US-0095486.
PR 06-AUG-1998; 98US-0095454.
PR 06-AUG-1998; 98US-0095455.
PR 12-AUG-1998; 98US-0096319.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Komatsu G, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Edner R, Olsen HS, Mucenski M;
XX
XX WPI; 2000-195282/17.
XX
XX New isolated human genes and the secreted polypeptides they encode,
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Disclosure: Page 18; 634p; English.
XX
XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
XX human secreted proteins given in AA191451 to AA191691. The human secreted
XX proteins can have activities based on the tissues and cells they are
XX expressed in. Examples of the activities are: cytostatic;
XX immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;
XX antiallergic; osteopathic; antiallergic; antibacterial; antidiabetic;
XX antitachy; antipruritic; and cardiant. The polynucleotides and their
XX corresponding secreted proteins are useful for preventing, treating or
XX ameliorating medical conditions, e.g. by protein or gene therapy. Also
XX pathological conditions can be diagnosed by determining the amount of the
XX proteins in a sample or by determining the presence of mutations in the
XX polynucleotides. Specific uses are described for each of the
XX polynucleotides, based on which tissues they are most highly expressed
XX in, and include developing products for the diagnosis or treatment of
XX cancer, tumours, neurodegenerative disorders, developmental abnormalities
XX and foetal deficiencies, blood disorders, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, inflammation,
XX allergies, Alzheimer's and behavioural disorders, schizophrenia,
XX osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
XX transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
XX cardiovascular disorders, reproductive disorders, gastrointestinal
XX disorders, respiratory disorders and metabolic disorders. The proteins
XX or polynucleotides can also be used as food additives or preservatives.
XX The proteins are also useful for identifying their binding partners.
XX AAA26337 to AAA26345 and AA191450 are sequences used in the
XX exemplification of the present invention.
XX
XX Sequence 443 AA:
SQ
Query Match 49.0%; Score 2155; DB 21; Length 443;
Best Local Similarity 100.0%; Pred. No. 4,1e-179;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAGLNCVSIALGVLLGARNLRPGAAFEIALPRESNITVLIKGTPTLLAKPCYIYI 60
|||||

Db 30 MAGLNCVSIALGVLLGARNLRPGAAFEIALPRESNITVLIKGTPTLLAKPCYIYI 89
OY 61 SKRHITMLSIKSGERIVFFPSCSPENHVEIEQKINIDCMGCPGEPQLOPSTSLP 120
|||||
Db 90 SKRHITMLSIKSGERIVFFPSCSPENHVEIEQKINIDCMGCPGEPQLOPSTSLP 149
OY 121 LNTFTIMDKAHSISLELOFSTIPRLROIQPGSCPDGYTHSISGRIDATVVAIGFCSN 180
150 LNTFTIMDKAHSISLELOFSTIPRLROIQPGSCPDGYTHSISGRIDATVVAIGFCSN 209
Db 181 GTVSRIKMOEGVKMALHPHFHPRNVSGFSIANRSSIKRLCIIEVEEGSATLMSANY 240
|||||
Db 210 GTVSRIKMOEGVKMALHPHFHPRNVSGFSIANRSSIKRLCIIEVEEGSATLMSANY 269
OY 241 PEGFPEDELMTQFVPAHLRASVFLNRLNSCEKKEVEYIIGSTTNPEVFLDEK 300
|||||
Db 270 PEGFPEDELMTQFVPAHLRASVFLNRLNSCEKKEVEYIIGSTTNPEVFLDEK 329
OY 301 OPGNMAGNFNLISLOGCDODASPGILRLQFVLYOHPONESNKIYVVDLSNERAMSLTTE 360
|||||
Db 330 OPGNMAGNFNLISLOGCDODASPGILRLQFVLYOHPONESNKIYVVDLSNERAMSLTTE 389
OY 361 PRPVKSRKRFVPCFVCLSRCTCSNLTLSGSKHKISFLCDDLTRLMNVK 413
|||||
Db 390 PRPVKSRKRFVPCFVCLSRCTCSNLTLSGSKHKISFLCDDLTRLMNVK 442
RESULT 6
AAU12252
ID AAU12252 standard; Protein; 343 AA.
XX
XX AAU12252:
AC
XX 24-OCT-2001 (first entry)
XX
XX Human PRO5773 polypeptide sequence.
DE
XX
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy.
OS Homo sapiens.
XX
XX WO200140466-A2.
PN
XX 07-JUN-2001.
PD
XX
XX 01-DEC-2000; 2000WO-US32678.
PF
XX
XX 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31243.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.

22-MAY-2000: 2000MO-US14042.
 30-MAY-2000: 2000MO-US14941.
 PR 02-JUN-2000: 2000MO-US15264.
 PR 10-NOV-2000: 2000MO-US30873.

XX (GETH) GENENTECH INC.

XX Baker KP, Beresini M, DeForge L, Desnoyers L, Ellvaroff E, Gao W;
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX WPI: 2001-408281/43.
 DR N-PSDB; AAS21324.

DR Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical -

PS Claim 12: Fig 162; 813pp; English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

SO Sequence 343 AA:

Query Match 40.5%; Score 1779; DB 22: Length 343;
 Best Local Similarity 99.4%; Pred. No. 1.8e-146;
 Matches 341; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGNGCVSIALGVLILGAARLPRGAEAFILPRESNTIVLTKICTPTLLAKPCYIV 60
 DB 1 MAGNGCVSIALGVLILGAARLPRGAEAFILPRESNTIVLTKICTPTLLAKPCYIV 60
 QY 61 SKRHITMLSTKSGERTVFTSCSPENHVEITOKNIDCSGPCPEGEVOLQPSSTLPT 120
 DB 61 SKRHITMLSTKSGERTVFTSCSPENHVEITOKNIDCSGPCPEGEVOLQPSSTLPT 120
 QY 121 LNTFTMDVKAHKSIGLELOFSTPLRLQIGPGSCDPGVTHSISGRIDATVVRIGTFCSN 180
 DB 121 LNTFTMDVKAHKSIGLELOFSTPLRLQIGPGSCDPGVTHSISGRIDATVVRIGTFCSN 180
 QY 181 GTSRITKMGCVKMAHLMPHFHRNVSQFSINARSSIKRCLCTIESVPEGGSATLMSAN 240
 DB 181 GTSRITKMGCVKMAHLMPHFHRNVSQFSINARSSIKRCLCTIESVPEGGSATLMSAN 240
 QY 241 PEGFPEDELMTQFVPAHLRASVSFLNPLNSCERKEEVEYYIGSTNPNPEFKLEDK 300
 DB 241 PEGFPEDELMTQFVPAHLRASVSFLNPLNSCERKEEVEYYIGSTNPNPEFKLEDK 300
 QY 301 QPQNMAGNENLSTLQCGDQDQSPGILRLQFVLVQHPRQMSNK 343
 DB 301 QPQNMAGNENLSTLQCGDQDQSPGILRLQFVLVQHPRQMSNK 343

RESULT 7
 AA91593
 ID AA91593 standard; Protein: 159 AA.
 XX
 AC AA91593;

DT 29-JUN-2000 (first entry)

DE Human secreted protein sequence encoded by gene 6 SEQ ID NO:266.

XX Human: secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antiinflammatory; neutropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antisthma;
 KW antipostic; cardiant; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour.

OS Homo sapiens.

PN WO200006698-A1.

PD 10-FEB-2000.

PF 29-JUL-1999; 99MO-US17130.

XX 30-JUL-1998; 98US-0094657.

PR 05-AUG-1998; 98US-0095486.

PR 06-AUG-1998; 98US-0095454.

PR 06-AUG-1998; 98US-0095455.

PR 12-AUG-1998; 98US-0096319.

XX (HUMA-) HUMAN GENOME SCI INC.
 XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenik M;
 XX WPI: 2000-195282/17.

DR New isolated human genes and the secreted polypeptides they encode,
 XX useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX
 PS Disclosure: Page 18; 634pp; English.

XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AA91451 to AA91691. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antiinflammatory; neutropic; neuroprotective;
 CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 CC antisthma; antipostic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The proteins
 CC or polynucleotides can also be used as food additives or preservatives.
 CC The proteins are also useful for identifying their binding partners.
 CC AAA26337 to AAA26345 and AA91450 are sequences used in the
 CC exemplification of the present invention.

Sequence 159 AA:

Query Match 18.8%; Score 827; DB 21; Length 159;
 Best Local Similarity 100.0%; Pred. No. 6-9e-64;
 Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 FEIALPRESNITVLKLGPTLLAKPCYIVISKRHITMISIKSGERIVTFSCOSPENHF 89
 DB 1 FEIALPRESNITVLKLGPTLLAKPCYIVISKRHITMISIKSGERIVTFSCOSPENHF 60
 QY 90 VIEIQNNICMSPCPFGEVQLOPSTSLPTLTKRTFMDYKAKHSIGLEIQSIPRLRQI 149
 DB 61 VIEIQNNICMSPCPFGEVQLOPSTSLPTLTKRTFMDYKAKHSIGLEIQSIPRLRQI 120
 QY 150 GPGEPCPDGVTSHISGRIDATVVRIGTFCSNGTVSRIM 188
 DB 121 GPGEPCPDGVTSHISGRIDATVVRIGTFCSNGTVSRIM 159

RESULT 8
 ABG00296
 ID ABG00296 standard; Protein; 143 AA.
 AC ABG00296;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #287.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YF;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS64483.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID No 30655; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 143 AA;
 QY Query Match 17.6%; Score 775; DB 22; Length 143;
 DB Best Local Similarity 100.0%; Pred. No. 2e-59;
 Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 694 KKKTKNGPAVGTYNGNINTEMPROPKFKGGRKNDSHYAVIEDTMVYGHLLDSSGS 753
 DB 1 KKKTKNGPAVGTYNGNINTEMPROPKFKGGRKNDSHYAVIEDTMVYGHLLDSSGS 60
 QY 754 FLOPEVDYRPFQGTMGVCPSPPTICSRAPTKLATEEPPSPSESEPYTFSHPN 813
 DB 61 FLOPEVDYRPFQGTMGVCPSPPTICSRAPTKLATEEPPSPSESEPYTFSHPN 120
 QY 814 GDVSSKDTDIPLLNTOEPMEPAE 836
 DB 121 GDVSSKDTDIPLLNTOEPMEPAE 143

RESULT 9
 AAM23969
 ID AAM23969 standard; Protein; 121 AA.
 AC AAM23969;
 DT 12-OCT-2001 (first entry)
 DE Human EST encoded protein SEQ ID NO: 1494.
 XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 XX biodiversity; gene therapy; nutrition.
 OS Homo sapiens.
 PN WO200154477-A2.
 PD 02-AUG-2001.
 PE 25-JAN-2001; 2001WO-US02687.
 PF 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX (HYSE-) HYSEQ INC.
 PA Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 DR WPI; 2001-476164/51.
 DR N-PSDB; AAH98628.
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 PS Claim 20; Page 1036; 1275pp; English.
 XX The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

Sequence 121 AA;
Query Match 14.0%; Score 616; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGINCVSIALGCVLLGAAARLPAGAEFIALPRESNTVLIKGTPTLLARPCYVI 60
D 1 MAGINCVSIALGCVLLGAAARLPAGAEFIALPRESNTVLIKGTPTLLARPCYVI 60
QY 61 SKRHTMSTKSGERIVFTSCSPENHFVTEIKNIDCMGPGCPGEVOLQPSSTLPT 120
D 61 SKRHTMSTKSGERIVFTSCSPENHFVTEIKNIDCMGPGCPGEVOLQPSSTLPT 120
QY 121 L 121
D 121 L 121

RESULT 10
AAU69483
ID AAU69483 standard; Protein: 167 AA.

AC AAU69483;
DT 29-JAN-2002 (first entry)
DE Human purified secretory polypeptide #52.
XX
KW Human; purified secretory polypeptide; cell proliferative disorder;
KW cancer; immune system disorder; neurological disorder; mental disorder;
KW motor neuron disorder; demyelinating disorder; neuromuscular disorder;
KW central nervous system disorder; enzyme linked immunosorbent assay;
KW ELISA; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200162918-A2.
XX
PD 30-AUG-2001.
XX
PE 01-FEB-2001; 2001WO-US03465.
XX
PR 24-FEB-2000; 2000US-185215P.
PR 24-FEB-2000; 2000US-185216P.
PR 16-MAY-2000; 2000US-205232P.
PR 17-MAY-2000; 2000US-205286P.
PR 17-MAY-2000; 2000US-205287P.
PR 17-MAY-2000; 2000US-205323P.
PR 17-MAY-2000; 2000US-205324P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'sa SA, Amshay S, Dahl CR, Dam JC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenawalt IB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE;
XX
DR WPI; 2001-648217/4.
XX
XX Nucleic acids encoding secretory polypeptides, useful in genetic
PT diagnosis and therapy -
PT
XX
PS Disclosure; Page 221; 237pp; English.
XX
CC Sequences AAU69432-AAU69511 represent purified secretory polypeptides of
CC the invention. The polypeptides and their associated polynucleotides can
CC be used in the treatment, prevention and diagnosis of diseases associated
CC with inappropriate secretory protein expression. These diseases include
CC proliferative disorders such as atherosclerosis and psoriasis,
CC as well as leukemia and melanoma, immune system disorders such as

CC asthma and diabetes mellitus, neurological disorders such as epilepsy and
CC Parkinson's disease, mental disorders such as schizophrenia and seasonal
CC affective disorder (SAD), motor neuron disorders such as amyotrophic
CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,
CC central nervous system disorders such as mental retardation and
CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and
CC muscular dystrophy. Target polynucleotides in a sample can be detected by
CC hybridizing the sample with a probe sequence complementary to the target
CC polynucleotide, under conditions in which a hybridisation complex is
CC formed, and detecting the presence or absence of the complex. The
CC polypeptides may also be used as antigens in the production of antibodies
CC against secretory proteins and in assays to identify modulators of
CC protein expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the sequences of the
CC invention in samples e.g. by enzyme linked immunosorbent assay (ELISA).
XX
SQ Sequence 167 AA;
Query Match 11.9%; Score 523; DB 22; Length 167;
Best Local Similarity 73.4%; Pred. No. 2.6e-37;
Matches 113; Conservative 7; Mismatches 32; Indels 2; Gaps 2;

QY 94 OKNIDCMGPGCP-EGEVLQPSSTLPTLNRTFTMDYKAHKSIGLEOFSTPRRLQIGP 152
D 9 RKNIELYVRPMSLGEVOLQPSSTLPTLNRTFTMDYKAHKSIGLEOFSTPRRLQIGP 68
QY 153 ESCPDGYTHSISGRIDATVVRIGTSCNGTVSRIRKMGCVKMLHPV-FHPRNVSGFSI 211
D 69 KSCPDDGYTHSISGRIDATVVRIGTSCNGTVSRIRKMGCVKMLHPV-FHPRNVSGFSI 128
QY 212 ANRSIRKLTIEVFEGEGSATLMSANYPCGP 245
D 129 ETRSSIRKLTIEVFEGEGSATLNVCOLPYKAP 162

RESULT 11
AAM90342
ID AAM90342 standard; Protein: 116 AA.
XX
AC AAM90342;
XX
DT 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen SEQ ID NO:17935.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 31-JAN-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.

CC Cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX

SQ Sequence 116 AA;

Query Match 8.98; Score 390; DB 22; Length 116;
 Best Local Similarity 98.7%; Pred. No. 5.9e-26;
 Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 341 SNKTYVDLSNERAMSLTIEPRPVKSKRFPGCFVCLSERTSSNLTITSGSKHKXISFL 400
 12 SDRITYVDLSNERAMSLTIEPRPVKSKRFPGCFVCLSERTSSNLTITSGSKHKXISFL 71

OY 401 CDDLTFLMNMVEKRTIS 416
 72 CDDLTFLMNMVEKRTIS 87

DB

RESULT 12
 AAY27020
 ID AAY27020 standard; Protein; 3623 AA.

XX AAY27020;

XX 08-OCT-1999 (first entry)

XX Rat cubilin protein sequence.

XX Cubilin; epithelial glycoprotein receptor; vitamin B12; endocytosis;
 XX toxicity; kidney; spleen; brain; liver; heart; thyroid; abortion; rat;
 XX proteinuria; fetal malformation; fetal development; kidney damage.

XX Rattus sp.

XX Key Location/Qualifiers

FT Peptide 1..20 /note- "endoplasmatic import signal sequence"

FT Protein 20 /note- "mature protein"

FT Modified-site 95 /note- "potential N-glycosylation site"

FT Modified-site 428 /note- "potential N-glycosylation site"

FT Modified-site 491 /note- "potential N-glycosylation site"

FT Modified-site 711 /note- "potential N-glycosylation site"

FT Modified-site 781 /note- "potential N-glycosylation site"

FT Modified-site 857 /note- "potential N-glycosylation site"

FT Modified-site 957 /note- "potential N-glycosylation site"

FT Modified-site 984 /note- "potential N-glycosylation site"

FT Modified-site 1168 /note- "potential N-glycosylation site"

FT Modified-site 1285 /note- "potential N-glycosylation site"

FT Modified-site 1307 /note- "potential N-glycosylation site"

FT Modified-site 1319 /note- "potential N-glycosylation site"

FT Modified-site 1332 /note- "potential N-glycosylation site"

FT Modified-site 1500 /note- "potential N-glycosylation site"

FT Modified-site 1551 /note- "potential N-glycosylation site"

FT Modified-site 1646 /note- "potential N-glycosylation site"

FT Modified-site 1671 /note- "potential N-glycosylation site"
 FT Modified-site 1802 /note- "potential N-glycosylation site"
 FT Modified-site 1819 /note- "potential N-glycosylation site"
 FT Modified-site 2085 /note- "potential N-glycosylation site"
 FT Modified-site 2117 /note- "potential N-glycosylation site"
 FT Modified-site 2274 /note- "potential N-glycosylation site"
 FT Modified-site 2400 /note- "potential N-glycosylation site"
 FT Modified-site 2531 /note- "potential N-glycosylation site"
 FT Modified-site 2581 /note- "potential N-glycosylation site"
 FT Modified-site 2610 /note- "potential N-glycosylation site"
 FT Modified-site 2813 /note- "potential N-glycosylation site"
 FT Modified-site 2875 /note- "potential N-glycosylation site"
 FT Modified-site 2945 /note- "potential N-glycosylation site"
 FT Modified-site 2989 /note- "potential N-glycosylation site"
 FT Modified-site 3042 /note- "potential N-glycosylation site"
 FT Modified-site 3106 /note- "potential N-glycosylation site"
 FT Modified-site 3125 /note- "potential N-glycosylation site"
 FT Modified-site 3165 /note- "potential N-glycosylation site"
 FT Modified-site 3268 /note- "potential N-glycosylation site"
 FT Modified-site 3283 /note- "potential N-glycosylation site"
 FT Modified-site 3290 /note- "potential N-glycosylation site"
 FT Modified-site 3357 /note- "potential N-glycosylation site"
 FT Modified-site 3400 /note- "potential N-glycosylation site"
 FT Modified-site 3430 /note- "potential N-glycosylation site"
 FT Modified-site 3533 /note- "potential N-glycosylation site"
 FT Misc-difference 37 /note- "potential N-glycosylation site"
 FT Misc-difference 80 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 108 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 148 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 224 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 234 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 235 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"
 FT Misc-difference 241 /note- "the His residue at this position is given as a Gly residue in the formal Seq ID listing"

FT Misc-difference 274 Gly residue in the formal Seq ID listing"
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 353
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 358
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 402
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 488
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 496
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 513
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 535
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 558
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 567
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 640
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 656
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 683
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 685
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 696
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 730
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 752
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 763
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 771
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 948
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 961
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 982
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 1032
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT Misc-difference 1089
FT /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"

FT Misc-difference 1135 /note="the His residue at this position is given as a
FT Gly residue in the formal Seq ID listing"
FT
Query Match
Best Local Similarity 18.8%; Pred. No. 0.07%;
Matches 150; Conservative 78; Mismatches 283; Indels 289; Gaps 33;
OY 70 IKSGERIVTFPCQSPENHFIIEQKNDICMSGPCFGEVQLOPSTS--LLPTLNFRTI- 126
DB 2306 VSSRERITVLFKFTDGGSSMGFAKXSIASCGETVS-GDSGYIESIGYPTLPYANNVFCO 2364
OY 127 MPYKAKHSIGLEQFIPRLQIQGEGSCPDGT-----HSISGRIDATVVRIGTCSCNGT 182
DB 2365 WFTRLGLPGHYLTLSFEDFNL-GSSPG--CTKDFEITWENTHSORV-----LERYCGNST 2415
OY 183 VSRKIMQEGVKMALHLPFHPRNVGSFISANRSSIKRLCIIESVGE--GSATLMSAN 239
DB 2416 PSSVDYSSNV-ASYKRVTDGVSATSGFRLOPKSS-RQVC-----GGDLHGPTGFTSPN 2467
OY 240 YPEGFPDELMWQFVVPAPHLRASVSFLNPNLSN-----C 274
DB 2468 YPNPNPHARICEWTITVQEGRRIVFTMLRSTOPOCNSHLLYENGIRSNPILQKLC 2527
OY 275 ERKEEVEYYIGESTNPEVFKLEKOP-----GNMAGNFNLSLQCDQDAQS 322
DB 2528 SRVNTNTERKSSGNIMK-VVFTDGSRPYGFTASTSTEDAVCGFLPSVSG--GNFSS 2584
OY 323 PGI-----LRLOFOVLVOHPQNESKRIYVVDLSNERNASLTIEPRPVKSRKFPVG 373
DB 2585 PGVNGIRDAYARNLDCEMTISNPRENSISITVELISTSHOCT----- 2629
OY 374 CFYCLESFRCSSNLITLSSSKHKISFLCDDLRLRMNVNKTISCTDHRCCQKRSYLQVP 433
DB 2630 -FDVLEFRVGDADGPL-----IEK-----PC-----SLAP 2654
OY 434 SDILHLP-----VELHDF----- 446
DB 2655 TAPLVIPYQVWIRREVSNERVEYTGFIEXSPTDCGCGITGNGVSSPNRYPLYSAMTH 2714
OY 447 -SWKLLVPRDRLSLVLPAAKIQLOHTEKPCNTFSYLVASAIPSDLYFGSPCGSGIR 505
DB 2715 CSMILKAPBGHTTLTSLDFLEAH----PTCSDSVTYVRNODSPSGYIGRCGGSVPR 2770
OY 506 QIQOVKON-ISVTLKTPAPSFQDEASRQGLTVSFIPIFKEE-----GVFTVPTDKXYVL 559
DB 2771 PIOSGSNOIIVFNT-----NNQGTGRFYATMTTNALCCGGTFHSANGT-----I 2816
OY 560 RPNMMDRGLPSLTSVSM-----NISVPRDOYACLFERKRSQ----- 596
DB 2817 KSPHMPQTFPENSRCSWTYITHDSKHWETISFDSNFRIPSSDSCONSFVKWGGRLMIK 2876
OY 597 -----VVCOTGRAFMIIQEOIRAREIPS----- 620
DB 2877 TLATSCGVAPSPITYTSNIGFIAVFGSEMAAQAQSASIPISNCGTFMTSPDITSPNF 2936
OY 621 -----LDBD-----VLKPSFH-----HSHWVAINSCSPISGQ 650
DB 2937 PKOYDNNMCTYLIDADPOLSVLITFVSFHLEDRSATGTCDDHGLHI- IKGNLISSTPL 2995
OY 651 LDLEFSVTLPRTVDLTVIL 670
DB 2996 VTIQSEETLPRLVGDGPIVL 3015
RESULT 13
AAB70070
ID AAB70070 standard; Protein; 595 AA.
XX
AC AAB70070;
XX
DT 14-MAY-2001 (first entry)
XX

PS Claim 1; Page 2077-2079; 2102pp; English.

CC The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin
 CC fusion proteins of the invention.

CC Sequence 595 AA:

Query Match 2.9%; Score 126; DB 23; Length 595;

Best Local Similarity 19.9%; Pred. No. 0.085;

Matches 135; Conservative 86; Mismatches 247; Indels 212; Gaps 34;

QY 225 SYFEEBGSKTLMSANYPPECFPEDELTMOQVYPAR---LRASVSFLNENLNCERKERY 281
 DB 33 TVQEGICVSVLCFSFYQPQN-----GWTASDPVHGYPFRAG-----DHV 70
 QY 282 EYVIGSTNPPEVFLKEDQPGNMAGNPLSLQCGDQDQSGILRLQFOVLYQHPQNS 341
 DB 71 SRNITVATNNPARVAOEE-----TRDRPH-LLGDQNNKD 103
 QY 342 NKIYVVDLSNERAMSLTEPRPVKOSRKFPQCFV-CLESRTCSNLTLSGSKHISPL 400
 DB 104 CTL-----SIRDTRESDACTVFCVGRGMKNKYYKQDLSVN-VTMS 144
 QY 401 CDDLTRLMNKNVEKITSCTDHRVCQRKSYSLQVPSDILH-----LPV-----ELHDF 446
 DB 145 QDLISRYRLVEPESYTV-----QEGLCVSPCSVLYPHYNMWTASSPYGSMFKGADI 197
 QY 447 SKLLVPRDRSLVLPACKLQOHN-----EKPCSTSYLVASAIIPDOLYFGSF 498
 DB 138 PVDIIPVANT-----TPSGKVOEDTHGRPLLLGDPPQTNCSLSINDARKDGSKYFQ-- 249
 QY 499 CPGGSIK--QIQVKNISVTLRTFAPSFQ-----QEASRQGLTVSFIPYFKEEG----- 545
 DB 250 VERGSRKNMYIYDKLSVHTALTLMPTESITPGTLESQHRLNLTCS-VPMACEQGTPTTT 308
 QY 546 -----VFVTPD-TKSKVYLRTPN-WDRGLPSLT-----SVSNISVPRDQ 584
 DB 309 WMGASVSSLDPTITRSSMLSLIPQPDHG-TSLTCQVTLPGAGVTMTAVRLNISVPQN 367
 QY 585 VACLTFEKE-RSGVVCQGRAFMIIOEQRTAEIFSLDED----- 624
 DB 368 LMTVFOGGGTASTLRNLSALSLEGGSLH--VCADVSNPPARLSWTGSLTSPSQS 425
 QY 625 ---VLPKPSFA---HSEFWNISNCSPTSGKOLDLFFSV--TLTPRTVDLVIIAAGV 675
 DB 426 SMLGELPRVHVKBEGETCAQN--PLGSHISLSLSLQMEYTGKMPISGIVLGAAG 483
 QY 676 GG---VLLISALGLITCCVKKKKKTKNGPAVGIVNGNINTEMPROKKFKQGRKDNDSH 732
 DB 484 GAGATALVELYFCIIFVVVRSCKKKSAR-PAVGV--GDTGME----- 522
 QY 733 VYAVIEDPMVYGHLODSSGFLQPEVDYTRPFGTGMGCPSPSPITCSRAPATKATEE 792
 DB 523 -----DANAARV---SASGRLIESPADSPPHNAPALATPSPE--GEIYASLSFK 572
 QY 793 PPRSPPESESEPTFSHPN 812
 DB 573 ARPYQFQEGEALGTERSEIN 592

RESULT 15

AAE20788
 ID AAE20788 standard; Protein; 3095 AA.

XX AAE20788;

AC 01-JUL-2002 (first entry)

DT Rat C3b/C4b complement receptor like protein.

DE Rat C3b/C4b complement receptor like protein.

KW Rat; C3b/C4b complement receptor-like molecule; immune system disorder;

KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;

KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;

KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;

KW transplant rejection; autoimmune disease; ischaemic condition; nocturnal;

KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;

KW infertility; vasodilator; obesity; cardiant.

OS Rattus rattus.

XX Key Location/Qualifiers

XX Misc-difference 196

XX /Label= Unknown

XX /note= "Xaa can be any amino acid"

XX /Label= Unknown

XX /note= "Xaa can be any amino acid"

XX W0200210199-A2.

XX 07-FEB-2002.

XX 24-JUL-2001; 2001WO-US23232.

XX 02-AUG-2000; 2000US-222504P.

XX 28-NOV-2000; 2000US-0728787.

XX (AMGE-) AMGEN INC.

XX Welcher AA, Elliott GS;

XX WPT. 2002-303934/34.

XX N-PSDB; AAD33319.

XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic

XX acid molecule, useful for treating, preventing and diagnosing

XX rheumatoid arthritis, psoriatic arthritis, inflammatory arthritis, and

XX multiple sclerosis -

XX Claim 13; Fig 3; 251pp; English.

XX The invention relates to a nucleic acid encoding a novel C3b/C4b

XX complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like

XX polypeptide and nucleic acid molecules may be used to treat, prevent,

XX ameliorate, diagnose and/or detect diseases such as immune system

XX disorders such as rheumatoid arthritis, psoriatic arthritis,

XX inflammatory arthritis, osteoarthritis, inflammatory joint disease,

XX autoimmune disease, multiple sclerosis, lupus, inflammatory bowel

XX disease, transplant rejection, nervous system disorders (e.g. Alzheimer's

XX disease), ischaemic conditions, metabolic disorders (e.g. obesity and

XX diabetes) and infertility. The invention is useful in gene therapy. The

XX present sequence is rat C3b/C4b complement receptor like protein.

XX Sequence 3095 AA:

Query Match 2.7%; Score 119; DB 23; Length 3095;

Best Local Similarity 17.8%; Pred. No. 4.6;

Matches 175; Conservative 89; Mismatches 296; Indels 422; Gaps 47;

QY 1 MAGLNGVSIALL---GVLLGAARLRGAEAFET---ALPRESNTVYIKIGTPIILAK 54

DB 1225 LSGSHGETPLATNSQILLRPSAKSGASAGARFHEVYQAVPTSD-----T 1270

QY 55 PCYIVSKRHITMLSIKSGRIVTFPS-----COSPENHVIYE 92

Db 1271 QCSSVPEPRY-----GRIGSEFSAGSIVRECPNGYLLQGSTAIRCOQVFN-ALAQ 1321
QY 93 IQKNIDCMGCPGEGEYQLOPSTSLPTL-----NRTFIWDVKAHKSIGLELOFSIPRL 146
Db 1322 WNDITIPSCVYPCS-GNFTQRRGTILSPGYEPYGNLNCWKITVSESGIQIOVISFAT 1380
QY 147 RQIGPGSCPDGVTHSISGRIDATVRIGTFCSSNGTVSRIKMOGVKMLHL----- 198
Db 1381 EQMNDSEIHGG-----DMTAPRLGSF-SGTIVBALLNSTSNOLCHFOQDISVAA 1431
QY 199 PMFH-----PRN-----VSGFSTANRST----- 217
Db 1432 AGFHLEKYTVGLAACQEPALPSNGIKIGDRYMNVDVLSFOCEPGYTLQGRSHISCMPTV 1491
QY 218 -----KRLCI-----IESVEEGESATLMSANYPEGPEDELMTOFVPAHLRASVS 265
Db 1492 RRMVYPSPLCIATCGGTLTSM-----SGVILSPGFGSYRNLDCTWKISLPIGYGAHQ 1546
QY 266 FLNFNLSCERKEERVE-----YY-----IPG---STNPEVFKLEDKQPGM 305
Db 1547 FLNF-----STEANHDYLEIQNGPYHSPMGQFSGPDLPASLSTHETLIRFYSDHSQNR 1603
QY 306 AGNFNLS-----IQGC-DQDAQSPGIL-----RLQFO-----VLVOHPONESKTY 345
Db 1604 QG-FKLSYQAYELQNCDDPPAFQNGFMINDSVGQISFECPGYILLGHP----- 1654
QY 346 VVDLSNERAMSLTIEPRPVKQSRKFVPGCFYCLSEFRTCSSNLTLTSGSKHISF----- 399
Db 1655 -----VLTCQHTDNNMNPFRPC-----DAPCGYNTSQNGTIYSPGPPDEYPI 1699
QY 400 -----LCDDLTRLW-----MNYEKTISCTDHRVCO 424
Db 1700 LKDCMLVTPVPGHGYINFTLLQTEAVNDYIAVMDGPDQNSPOLGVFSGNTALETAYSS 1759
QY 425 RKSYSLOVPDILH--LPELHDFSWKLLVPRDLSLVLPRAKLOQHTHE----- 473
Db 1760 TNOVLLFHSDFSGGFEVLFNFAFQKRCPPP-----VVPQADLLEDEDEFEIGDFVK 1814
QY 474 KPCNTSFSYL-----VASAIPSQDLYFGS-----FCPGGSIK----- 505
Db 1815 YQCHPGYTLGSDTLTKLSQLLFQGSPTCEAQCPANEVRETSQVILSPGPGNFTN 1874
QY 506 -----OIOVKQNTSVTLRTFAPSFOQE----- 527
Db 1875 SQTCAMSIKYEPMFNITL--FVDTFQSEKQFDALVFDGSSGSPLLVLSGNHTEQSNF 1932
QY 528 -----ASRQGLTVSF-IPY-----FKEGVFTVPTDF--KSKYTL----- 559
Db 1933 TSRSNHLYLKMSIDHATSKKGFKIRYAPYCSLTSTLKNGVLNKTAGALGSKVQYFCKP 1992
QY 560 -----RTP-----NMDRGLPSLTSVSWNISVPRQVACLTFFKERSGVVCOTG 602
Db 1993 GYRMIGHSNATCRNRPYGVYQWDSMAPLCAVSCGT-----PEAPGNGSFTG 2039
QY 603 RAEMIIQEQRTRAEELFSLD-----EDVL-----PKPSFH-----HHS 635
Db 2040 NEFTLDSKYVEECNEGFKLDASQOATFVQEDGLMSNRGKPPCKPVPSPSIEGOLSEHY 2099
QY 636 FMYNIS-----NCSF 645
Db 2100 LMRIVSGSLNEYGAVLLSCSP 2121

Search completed: November 11, 2002, 11:32:48
Job time : 72 secs

5387 ACCATCACCAGCATCATTGTTGCTTCGACAGSACACCACTGGCCATTTTCCCTTC-AACTG 5445

Db	519	ACNATCAGGAAATNNCTGTGCTGGCANGAGNCCAAAGGGGCAATTTCTTCAAAANTA	460
QY	5446	AGGGCTCAAAACCTCTGGACAAATTCGTGGCTC-TGACACCAGTATTTCTTGGAGCTGT	5504
Db	459	AGGGCTCAAAACCTCTGGACAAATTCGTGGCTCCTTTAAACNCCAGTATTTCTTGAAGCTGG	400
QY	5505	GCCCTAGTAAGGGGCCACAGCTGAGGAAACCCGCGCTCTTTTCTTTAAAGCCAGAGCCC	5564
Db	399	CCTCAAGTAAAGGGGCCACAGCTGAGGAAACCCGCGCTCTTTTCTTTAAAGCCAGAGCCC	340
QY	5565	ACTTACATAAACATTTTCAGGGTCACTGGAAACAGTGAAGTGCATTTGTTGAGGCTTAC	5624
Db	339	ACTTACATAAACATTTTCAGGGTCACTGGAAACAGTGAAGTGCATTTGTTTAAACCTTAC	280
QY	5625	TGCATGCGACGCCACAGCTCATATCCAGTGGTGTCGCATCGACGAAAGGCCAGCGCA	5684
Db	279	TGCATGCGACGCCACAGCTCATATCCAGTGGTGTCGCATCGACGAAAGGCCAGCGCA	220
QY	5685	TGCAGAGCTGTCTTAATGCTGTGGTTCATTGACAGAAAGGAAAGGTCTCAAGAGNAG	5744
Db	219	TGCAGAGCTGTCTTAATGCTGTGGTTCATTGACAGAAAGGAAAGGTCTCAAGAGNAG	160
QY	5745	TCAACTGGGACAAGCAACGCCACCGGACATCGCTTGGTAAAGTTAGCAGACTGGG	5804
Db	159	TCAACTGGGACAAGCAACGCCACCGGACATCGCTTGGTAAAGTTAGCAGACTGGG	100
QY	5805	TGTGTGATCTGCAGTGTCTTCACTGGAATAATTTATTCATTGCAGATCTATTATAGTG	5864
Db	99	TGTGTGATCTGCAGTGTCTTCACTGGAATAATTTATTCATTGCAGATCTATTATAGTG	40
QY	5865	GCATTTATTCATTTCTCTGTGCTTTAAATTAACAAATGT	5903
Db	39	GCATTTATTCATTTCTCTGTGCTTTAAATTAACAAATGT	1

```

RESULT 2
US-09-798-096-10/c
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Walt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

Query Match      3.5%; Score 218.6; DB 4; Length 99500;
Best Local Similarity 82.6%; Pred. No. 1.3e-41;
Matches 262; Conservative 0; Mismatches 54; Indels 1; Gaps 1.

OY 4492 TCCTTTTCTTTTTCTTTTTTTTTTAATGTGAGACAGAGATCATTCGTGGCCTAGGCT 4551
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66182 TTTTGTGTTTGTGTTTGTGTTTGTGTTGTTTGAGACAGAGTCTCACTGTGCCCCAGGCT 66123

OY 4552 GGAAGTCAGATGGCGCAATCTGGGCTCACTGCACACTCTGCTCTGGGGCTCAGAATAATC 4611
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66122 GGAATGTCAGTGGCACAAATCTTGGCTCACTACAACCTCTGCGTGGGTTCAAGCCATTTC 66063

OY 4612 TGCCACGTCACCCTCCCAAATAGCTGGGATCACTGSGCAACAAACCAACCATGCCAGACTAA- 4670
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66062 TCCGCTCAGCCCTCCCAAAGTAGCTGGGATTATGATGCCCAACCATCGCGCCAGCTAAT 66003

OY 4671 TTTTGTATTTTTTTNAGAGAGAGGGTTTCACCATGTTGCCAGAGGCTGTCACCACTCCT 4730
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 66002 TTTTGTATTTTTTAGAGACAGAGATTTTCACCATGTTGGCCAGGCTGGGTCTTCGAACCTCT 65943

```

QY	4731	GGGCTAAGCAATCCGCTCCGCTGGCTCCCAAGTCGTGGATTACACATTTGAGCA	4790
Db	65942	GACCTCAAGCAATCCGCTTCGCTTGGCTCCCAAGTCTGTGGATTACAGCATTTAGCCA	65883
QY	4791	CCGCAATCGAAGCCCAACA	4807
Db	65882	CCGGGCCCAAGCCCTACCA	65866

```

RESULT 3
US-08-646-301A-1/C
: Sequence 1, Application US/08646301A
: Patent No. 6194211
: GENERAL INFORMATION:
: APPLICANT: Richards, Cynthia Ann
: APPLICANT: Huber, Brian E.
: TITLE OF INVENTION: Transcriptional Regulatory Sequence of Carcinoembryonic
: Patent No. 6194211
: TITLE OF INVENTION: Antigen for Expression Targeting
: FILE REFERENCE: PBI508USM
: CURRENT APPLICATION NUMBER: US/08/646,301A
: CURRENT FILING DATE: 1996-05-16
: NUMBER OF SEQ. ID NOS: 25
: SOFTWARE: Patentln Ver. 2.1
: SEQ. ID NO 1
: LENGTH: 11288
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-646-301A-1

```

Query Match	3.5%;	Score 218.2;	DB 4;	Length 11286;
Best Local Similarity	81.7%;	Pred. No. 5.8e-42;		
Matches 264;	Conservative 0;	Mismatches 58;	Indels 1;	Gaps 1;

QY 4480 TTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTTATGTGAGACAGCATCTCATTC 4539
 Db 858 TTTTCTTTTCTTTCTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 799
 QY 4540 GTTCCATAGGCTGGAGTGCAGTGGGGCAATCTGGCTACTGCACACCTGGCTCTGG 4599
 Db 798 GTTCCCAAGCTGGAGTGCAGTGGGCAATCTGGCTACTGCACACCTGGCTCTGG 739
 QY 4600 CTCAGCAATTTCTCCCACTCAGCTCCCAATATGCTGGATCATCTGGCACAAACACCA 4659
 Db 738 TCCAGCAATTTCTCAGCTCAGCTCCCAATATGCTGGATCATCTGGCACAAACACCA 679
 QY 4660 TGCCAGCTAA-TTTTGTATTTTGTAGAGACAGGGTTTACCATTGTTGCCAGAGCTGG 4718
 Db 678 TGCCAGCTAATTTTGTATTTTGTAGAGACGGGGTTTACCATTATTTGGCAGAGTGG 619
 QY 4719 TCTCAACCTCTGGGCTTCAGCAATCTCTCTGCTCTGGCTTCCCAAGTCTGGGATTTC 4778
 Db 618 TCTTGAACCTCTGACCTTCAGCAATCTGCCCCACCTCAACCTTCTTAACTGCTGGGATTTC 559
 QY 4779 AGATGTAGAGCCAGCATCCAGC 4801
 Db 558 AGCGGTAGCCAGCGCCGCGC 536
 RESULT 4
 US-08-481-968A-4/c
 Sequence 4, Application US/08481968A
 Patent No. 6300490
 GENERAL INFORMATION
 APPLICANT: Huber, Brian
 APPLICANT: Richards, Cynthia
 TITLE OF INVENTION: Molecular Constructs Comprising a Carcinoembryonic Antigen (CEA) Polypeptide
 FILE REFERENCE: PB1087US4
 CURRENT APPLICATION NUMBER: US/08/481,968A
 CURRENT FILING DATE: 1998-06-07
 NUMBER OF SEQ ID NOS: 36

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 11288
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-481-968A-4

```

Query Match	3.5%	Score 218.2;	DB 4;	Length 11288;
Best Local Similarity	81.7%;	Pred. NO. 5.8e-42;		
Matches 264; Conservative	0;	Mismatches 58;	Indels 1;	Gaps 1;

[illegible]

```

RESULT 5
US-08-154-712B-4/C
: Sequence 4, Application US/08154712B
: Patent No. 6337209
: GENERAL INFORMATION:
: APPLICANT: Huber, Brian
: APPLICANT: Richards, Cynthia
: TITLE OF INVENTION: Molecular Constructs Containing a Carcinoembryonic Antigen Reg
: TITLE OF INVENTION: Sequence
: FILE REFERENCE: PBI087053
: CURRENT APPLICATION NUMBER: US/08/154,712B
: CURRENT FILING DATE: 1993-11-19
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 11288
: TYPE: DNA
: ORGANISM: Homo sapiens
US-08-154-712B-4

```

[illegible][illegible]

```

: RESULT 6
: US-09-167-681-45
: Sequence 45, Application US/09167681A
: Patent No. 6265561
: GENERAL INFORMATION:
: APPLICANT: Welshillbom, M.D., Richard M.
: APPLICANT: Ratogianis, Rebecca B.
: APPLICANT: Wood, Thomas C.
: APPLICANT: Olterness, Diane M.
: TITLE OF INVENTION: SULFOTRANSFERASE SEQUENCE VARIANTS
: FILE REFERENCE: 07039/118001
: CURRENT APPLICATION NUMBER: US/09/167,681A
: CURRENT FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 52
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 45
:

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (4361)...(4507)
? NAME/KEY: CDS
? LOCATION: (4612)...(4737)
? NAME/KEY: CDS
? LOCATION: (4827)...(4925)
? NAME/KEY: CDS
? LOCATION: (6332)...(6447)
? NAME/KEY: CDS
? LOCATION: (6543)...(6638)
? NAME/KEY: CDS
? LOCATION: (7137)...(7316)
? NAME/KEY: CDS
? LOCATION: (7439)...(7553)
US-09-167-681-45

```

[illegible]


```

: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION: 2 kb PCR product using
: OTHER INFORMATION: primers, SEQ ID: 13 and 14; Also referred
: OTHER INFORMATION: to as JT108
: US-08-367-841a-11

```

Query Match 3.5%; Score 216.8; DB 4; Length 1988;
Best Local Similarity 82.8%; Pred. NO. 5.4e-42;
Matches 256; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

[illegible]

RESULT 13
 PCT-US95-07201-11
 Sequence 11, Application PC/TUS9507201
 GENERAL INFORMATION:
 APPLICANT: Chader, Gerald J.; Becerra, Sofia
 APPLICANT: Patricia; Schwartz, Joan P.;
 APPLICANT: Taniwail, Takayuki
 TITLE OF INVENTION: PIGMENT EPITHELIUM
 -TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
 ORGANIZATION AND SEQUENCE OF THE PDF GENB
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Morgan & Flinnegan, L.L.P.
 STREET: 345 Park Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/07201
 FILING DATE: 06-JUN-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/367,841
 FILING DATE: 30-DEC-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/257,963
 FILING DATE: 07-JUN-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/952,796
 FILING DATE: 24-SEP-1992
 ATTORNEY/AGENT INFORMATION:

```

NAME: DOROTHY R. AUTHER
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1988 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: JTB8A
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 2 kb PCR product using
OTHER INFORMATION: primers, SEQ ID: 13 and 14; Also referred
to as JT108
PCT-US95-07201-11

```

	Query Match	Similarity	3.5%	Score	216.8	DB	5	Length	1988
	Best Local	Similarity	82.8%	Pred.	5.4e-42				
	Matches	256	Conservative	0	Mismatches	52	Indels	1	Gaps
QY	4495	TTTTCTTTTTCTCTTTTTTTTTTAAATGTGAGACAGATCTCATTTCTGTTCCTAGGCTGGA	4554						
Db	532	TTTTTTTTTTTTTTTTTTTTNNNNCTTCTGAGAGGAGTCTCGCTTTGTCCACCAAGGTGGA	591						
QY	4555	GTGCAGTGGCGCAATCTCGGCTCACTGCACACTCTGCTCTCGGGCTCAAGCAATTCTCC	4614						
Db	592	GTCAGTGGTGTGATCTCAGCTCACTGCACAACTCTGCTCCCTCGTTCAAGGAATTTCTCC	651						
QY	4615	CACCCAGCGTCCCAATATGATGCGGGATCATCTGSCCAAAACCAACCAATGCCACACTAA	4673						
Db	652	TGCGTCACGCTCCAGAGTAGTGGGATTTACAGGACCTGCCATCAATGCCCACTAATTTTT	711						
QY	4674	TGTATTTTTTTGTAGAGACAGGGTTTCACCATTTGCCAGGCTGTCTCAACTCTCTGGG	4733						
Db	712	TGTATTTTTTAGTAGAGACGGGGTTTCACCATTTTGCTCTAGGCTGGGTCTAAACTCCGGG	771						
QY	4734	CTCAAGCAATCTCTCGCTCGGGCTCCCAAAAGTCTGGGATTACAGATGTAGCCACCG	4793						
Db	772	CTCAAGCATCTCACCGCGCTTGCTCCCAAAAGTCTGGGATTACAGGCTGAGCCACCG	831						
QY	4794	CATCCAGCC	4802						
Db	832	CGCTCGGCC	840						

RESULT 14
 US-08-520-373D-5
 : Sequence 5, Application US/08520373D
 : Patent No. 6451763
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Tombran-Tink, Joyce
 : APPLICANT: Steele, Flintan R
 : APPLICANT: Chader, Gerald J
 : APPLICANT: Becerra, Sofia P
 : APPLICANT: Johnson, Lincoln V
 : APPLICANT: Rodriguez, Ignacio R
 :
 : TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR
 : FILE REFERENCE: 2026-4203051
 :
 : CURRENT APPLICATION NUMBER: US/08/520,373D
 :
 : CURRENT FILING DATE: 1995-08-29
 :
 : PRIOR APPLICATION NUMBER: 08/377,710
 :
 : PRIOR FILING DATE: 1995-01-25
 :
 : PRIOR APPLICATION NUMBER: 08/279,979
 :
 : PRIOR FILING DATE: 1994-07-25
 :
 : PRIOR APPLICATION NUMBER: 07/894,215

```

; PRIOR FILING DATE: 1992-06-04
; PRIOR APPLICATION NUMBER: 07/952,796
; PRIOR FILING DATE: 1992-09-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 5
; LENGTH: 5262
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; OTHER INFORMATION: EXON 35-161; EXON 1142-1297; EXON 1984-2187;
; OTHER INFORMATION: EXON 5170-5255; INTRON 162-1141; INTRON
; OTHER INFORMATION: 1298-1983; INTRON 2188-5169; CDS 35-161; CDS
; OTHER INFORMATION: 1142-1297; CDS 1984-2187; CDS 5170-5255
; NAME/KEY: exon
; LOCATION: (35)..(160)
; NAME/KEY: exon
; LOCATION: (1142)..(1297)
; NAME/KEY: exon
; LOCATION: (1984)..(2187)
; NAME/KEY: exon
; LOCATION: (5170)..(5256)
; NAME/KEY: Intron
; LOCATION: (162)..(1141)
; NAME/KEY: Intron
; LOCATION: (1298)..(1983)
; NAME/KEY: Intron
; LOCATION: (2188)..(5169)
; OTHER INFORMATION: n = a or g or t or c, any base
US-08-520-373D-5

Query Match          3.5%; Score 216.8; DB 4; Length 5262;
Best Local Similarity 82.8%; Pred. No. 8.6e-42;
Matches 256; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4495 TTTCTTTTCTTTTCTTTTCTTTTATGTGAGACAGATCTCTCTGCTAGGCTGGA 4554
DB 564 TTTTCTTTTCTTTTCTTTTCTTTTNNCTTTCTGAGACGAGTCTGCTTGTCCCAAGCTGGA 623
QY 4555 GTGACGTGGGCAATCTCGGCTCACTGCAACCTCTGCTCTGCGGCTCAAGCAATTTCTCC 4614
DB 624 GTGACGTGGGATGATCTGATCTGATCTGCAACCTCTGCTCTGCGGCTCAAGCAATTTCTCC 683
QY 4615 CACCTCAGCTCCCAATAGCTGGGATCTGCGACAAACCAACCATGCTAA-TTT 4673
DB 684 TGCTTCAGCTCCAGATAGCTGGGATTTACAGGACCTGCTCATCATGCTAAATTTT 743
QY 4674 TGTATTTTGTAGAGACAGGCTTTCACANTGTGCCAGGCTGCTCAACCTCCCTGGG 4733
DB 744 TGTATTTTGTAGAGACAGGCTTTCACANTGTGCCAGGCTGCTCAACCTCCCTGGG 803
QY 4734 CTCAAGCAATCTCTCTCTGCGCTCCCAAGTGTGGGATTACAGATGTGAGCCACCG 4793
DB 804 CTCAAGGATCCACCGCCTTGCTCCCAAGTGTGGGATTACAGGCGTGAACCCACCG 863
QY 4794 CATCCAGCC 4802
DB 864 CGCCTGGCC 872

RESULT 15
US-08-367-841A-43
; Sequence 43, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tomdrian-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIAL
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Morgan & Flinnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: P1-147
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: full length genomic
; OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match          3.5%; Score 216.8; DB 4; Length 22481;
Best Local Similarity 82.8%; Pred. No. 1.7e-41;
Matches 256; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4495 TTTCTTTTCTTTTCTTTTCTTTTATGTGAGACAGATCTCTCTGCTAGGCTGGA 4554
DB 15129 TTTTCTTTTCTTTTCTTTTCTTTTNNCTTTCTGAGACGAGTCTGCTTGTCCCAAGCTGGA 15188
QY 4555 GTGACGTGGGCAATCTCGGCTCACTGCAACCTCTGCTCTGCGGCTCAAGCAATTTCTCC 4614
DB 15189 GTGACGTGGGATGATCTGATCTGATCTGCAACCTCTGCTCTGCGGCTCAAGCAATTTCTCC 15248
QY 4615 CACCTCAGCTCCCAATAGCTGGGATCTGCGACAAACCAACCATGCTAA-TTT 4673
DB 15249 TGCTTCAGCTCCAGATAGCTGGGATTTACAGGACCTGCTCATCATGCTAAATTTT 15308
QY 4674 TGTATTTTGTAGAGACAGGCTTTCACANTGTGCCAGGCTGCTCAACCTCCCTGGG 4733
DB 15309 TGTATTTTGTAGAGACAGGCTTTCACANTGTGCCAGGCTGCTCAACCTCCCTGGG 15368
QY 4734 CTCAAGCAATCTCTCTCTGCGCTCCCAAGTGTGGGATTACAGATGTGAGCCACCG 4793
DB 15369 CTCAAGGATCCACCGCCTTGCTCCCAAGTGTGGGATTACAGGCGTGAACCCACCG 15428
QY 4794 CATCCAGCC 4802
DB 15429 CGCCTGGCC 15437

Search completed: November 11, 2002, 11:36:24
```

Tue Nov 12 18:03:24 2002

Job time : 724 secs

us-09-899-569a-3.rni

This Page Blank (uspto)

|||||
Db 2101 ATCATCAGAGACGAGCGGAGCCGGGCTGAGGAGATCTCAGCCTGAGACGAGATGTCGTC 2160
Oy 2161 CCCAAGCCAAAGCTTCACCATCAGACGCTTCGGGGTAAACATCTCTACTGACGCCCCAGC 2220
Db 2161 CCCAAGCCAAAGCTTCACCATCAGACGCTTCGGGGTAAACATCTCTACTGACGCCCCAGC 2220
Oy 2221 ACGGCAAGCAGCTAGACCTGGCTTCGCGGTGACACTTACCCCAAGGAGCTGGGACTTG 2280
Db 2221 ACGGCAAGCAGCTAGACCTGGCTTCGCGGTGACACTTACCCCAAGGAGCTGGGACTTG 2280
Oy 2281 ACTGTCATCTCATGCGACGCGGTGGGAGGTGAGTCTTACTGCTGTCGCGCTCGGGCTC 2340
Db 2281 ACTGTCATCTCATGCGACGCGGTGGGAGGTGAGTCTTACTGCTGTCGCGCTCGGGCTC 2340
Oy 2341 ATCATTTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGAGGCCCGCGCTGTGGGTATC 2400
Db 2341 ATCATTTGCTGTGTGAAAAAGAAAGAAAGAACAAACAGAGGCCCGCGCTGTGGGTATC 2400
Oy 2401 TACAATGGCAATCAATACTGATGCGGAGGACGCCAAAAAGTTTCAGAAAGGGCGA 2460
Db 2401 TACAATGGCAATCAATACTGATGCGGAGGACGCCAAAAAGTTTCAGAAAGGGCGA 2460
Oy 2461 AAGGACAATGACTCCCATGTGTATGCACTCATGAGACACCATGGTATATGGGCACTTG 2520
Db 2461 AAGGACAATGACTCCCATGTGTATGCACTCATGAGACACCATGGTATATGGGCACTTG 2520
Oy 2521 CTACAGAGATTCCAGCGCGCTCTTCGACAGCCAGAGGTGGACACTACCGCGCGTTCCAG 2580
Db 2521 CTACAGAGATTCCAGCGCGCTCTTCGACAGCCAGAGGTGGACACTACCGCGCGTTCCAG 2580
Oy 2581 GGCACCATGGGGGTGTGCTCCTCCCTCCGACCCACCATATGTCGCGAGGGCCCCAACTGCA 2640
Db 2581 GGCACCATGGGGGTGTGCTCCTCCCTCCGACCCACCATATGTCGCGAGGGCCCCAACTGCA 2640
Oy 2641 AAGTTGGCCACTGAGAGGCGACCTCCTCGCTCCCTCTCTGAGTCTAGAGTGAACCGTAC 2700
Db 2641 AAGTTGGCCACTGAGAGGCGACCTCCTCGCTCCCTCTCTGAGTCTAGAGTGAACCGTAC 2700
Oy 2701 ACCTTCTCCCATCCCAACAATGGGGATGTAAAGCAGCAGACACAGACATTTCCCTTACTG 2760
Db 2701 ACCTTCTCCCATCCCAACAATGGGGATGTAAAGCAGCAGACACAGACATTTCCCTTACTG 2760
Oy 2761 AACACTCAGAGCGCATGAGGCGACGAGAAATACTGATCCACTTCCAGACGCTTTCCTCA 2820
Db 2761 AACACTCAGAGCGCATGAGGCGACGAGAAATACTGATCCACTTCCAGACGCTTTCCTCA 2820
Oy 2821 GTTTCATTAAGCAGGCGCACTGAGACACCGCTCGTCTTCTTAACAGAAATCCTTAAGAA 2880
Db 2821 GTTTCATTAAGCAGGCGCACTGAGACACCGCTCGTCTTCTTAACAGAAATCCTTAAGAA 2880
Oy 2881 GAGGAATTAATACAGAAAGAAACAGCAGAGAGTTTTCTGTGACACCGCCAACTTCACATTGC 2940
Db 2881 GAGGAATTAATACAGAAAGAAACAGCAGAGAGTTTTCTGTGACACCGCCAACTTCACATTGC 2940
Oy 2941 TAGATGGACTCATCTTAAGGGCAAGCATTTGAATAATGATGAATTCCAATCTGGATCACT 3000
Db 2941 TAGATGGACTCATCTTAAGGGCAAGCATTTGAATAATGATGAATTCCAATCTGGATCACT 3000
Oy 3001 CATGACAGCTCATGTCGCTCTCAACTTAAGCTGTGGGTAGGACAGCTGTAAATGAGAG 3060
Db 3001 CATGACAGCTCATGTCGCTCTCAACTTAAGCTGTGGGTAGGACAGCTGTAAATGAGAG 3060
Oy 3061 AGAGAGGCGCTGAGTCACTAGCATAGGGTTGACAGCAAGCCCTGAGATTACAGAGTGAAC 3120
Db 3061 AGAGAGGCGCTGAGTCACTAGCATAGGGTTGACAGCAAGCCCTGAGATTACAGAGTGAAC 3120
Oy 3121 AAGAGCTTCCCTTTCCAGAGACAAGTTCCAATTCCAAAGAGCCCTACCTGAGGTCCTCA 3180
Db 3121 AAGAGCTTCCCTTTCCAGAGACAAGTTCCAATTCCAAAGAGCCCTACCTGAGGTCCTCA 3180
Oy 3181 CTCTCACTGGGGTCCCAAGATGAAAAAGCAATGTCCTTTTATTTATTTATTTG 3240
Db 3181 CTCTCACTGGGGTCCCAAGATGAAAAAGCAATGTCCTTTTATTTATTTATTTG 3240

Db 3181 CTCTCACTGGGGTCCCAAGATGAAAAAGCAATGTCCTTTTATTTATTTATTTG 3240
Oy 3241 TGGTCTGTGTATTATTAAGAGATCAAAATGATTAACCACTAGCTCTTTTACCTGACTTA 3300
Db 3241 TGGTCTGTGTATTATTAAGAGATCAAAATGATTAACCACTAGCTCTTTTACCTGACTTA 3300
Oy 3301 GTAATTAACATCTTAACCTGTTTGGATGCTGGGTGTGACCTTTCACAGCCGCTAGAT 3360
Db 3301 GTAATTAACATCTTAACCTGTTTGGATGCTGGGTGTGACCTTTCACAGCCGCTAGAT 3360
Oy 3361 AAACGTGTGCTGTCCCGCAGGTGTGGGAATTAATTAATCAATCTGTCCAACGAAAAAGA 3420
Db 3361 AAACGTGTGCTGTCCCGCAGGTGTGGGAATTAATTAATCAATCTGTCCAACGAAAAAGA 3420
Oy 3421 ATGTGTGTGTTAGACAGATTGACACATATCTGCTTTGATTAAGAGACTTCGTATTC 3480
Db 3421 ATGTGTGTGTTAGACAGATTGACACATATCTGCTTTGATTAAGAGACTTCGTATTC 3480
Oy 3481 TAGTGGGTGCGTGGTATATCCATTTGTGAAATTCATCTTGATATCCCATTTGCTATAGT 3540
Db 3481 TAGTGGGTGCGTGGTATATCCATTTGTGAAATTCATCTTGATATCCCATTTGCTATAGT 3540
Oy 3541 CCTAGCAATTAAGAAATTTCTCAAGTTTCATGTGCGGGTCTCCTAGCTGACGAAATA 3600
Db 3541 CCTAGCAATTAAGAAATTTCTCAAGTTTCATGTGCGGGTCTCCTAGCTGACGAAATA 3600
Oy 3601 CTTTGACATTTAAAGAAATTTAGAGAAATATCTCATCTCTTAAAAATGTTTAAATATA 3660
Db 3601 CTTTGACATTTAAAGAAATTTAGAGAAATATCTCATCTCTTAAAAATGTTTAAATATA 3660
Oy 3661 TACCAAAAGTGGCGCCCGCTAGTTTCTGTGCACTGTGCAACCAATCTACTGTTAG 3720
Db 3661 TACCAAAAGTGGCGCCCGCTAGTTTCTGTGCACTGTGCAACCAATCTACTGTTAG 3720
Oy 3721 CTTAAAAACACATTAAGCTTATAGTCTCGGGGATCAGAAATTCAAAAATGATGTCCCT 3780
Db 3721 CTTAAAAACACATTAAGCTTATAGTCTCGGGGATCAGAAATTCAAAAATGATGTCCCT 3780
Oy 3781 GAATGAATTAAGAGTGTACAGAGAGCTGTCTCTTGTGAGGCTGTAGGAGAAAGCG 3840
Db 3781 GAATGAATTAAGAGTGTGTACAGAGAGCTGTCTCTTGTGAGGCTGTAGGAGAAAGCG 3840
Oy 3841 GTTCCTTGCATTTCAAGCTTCTAGAGGCGTGGATGCCAATGCCGCTCCAGTGGCTGTC 3900
Db 3841 GTTCCTTGCATTTCAAGCTTCTAGAGGCGTGGATGCCAATGCCGCTCCAGTGGCTGTC 3900
Oy 3901 AAGCTTTTCTCACATGGCACTGTGACACTGGCCCTCCCACTTCCCTTTGACTTAC 3960
Db 3901 AAGCTTTTCTCACATGGCACTGTGACACTGGCCCTCCCACTTCCCTTTGACTTAC 3960
Oy 3961 AAAGCCACAGGAAGATCCAGATTAATCTCTCATCTTAAGATCCTTCATCTGGA 4020
Db 3961 AAAGCCACAGGAAGATCCAGATTAATCTCTCATCTTAAGATCCTTCATCTGGA 4020
Oy 4021 AGAGCCTTTTGCATGCAAGACAAATATGCCAAGGTGGGATTAAGACAGCAATCTT 4080
Db 4021 AGAGCCTTTTGCATGCAAGACAAATATGCCAAGGTGGGATTAAGACAGCAATCTT 4080
Oy 4081 TGGGGTGTGTATTCTGTGCTTACACACCTTCTGTGCACACTGCTCCACAGAGAGCTTA 4140
Db 4081 TGGGGTGTGTATTCTGTGCTTACACACCTTCTGTGCACACTGCTCCACAGAGAGCTTA 4140
Oy 4141 CAAATGATCTGGGCAAGGATGTTTGTGTTAGCTTGGGACTTAACCTTAAAAAA 4200
Db 4141 CAAATGATCTGGGCAAGGATGTTTGTGTTAGCTTGGGACTTAACCTTAAAAAA 4200
Oy 4201 ACCCGATTCAGAAATCTGGCCATGCTGGGCTCATCTTCACCTAGCAACAATGGGC 4260
Db 4201 ACCCGATTCAGAAATCTGGCCATGCTGGGCTCATCTTCACCTAGCAACAATGGGC 4260
Oy 4261 TGGAGCTGGGCAACCACTGTGCTTTAGAAAGGGGTGTCACTTCCACAGGTCCACAGC 4320
Db 4261 TGGAGCTGGGCAACCACTGTGCTTTAGAAAGGGGTGTCCACTTCCACAGGTCCACAGC 4320

OY	43321	CCACGACATCGCCCATACACTTCCACAAATAGAGCTAAGTGTGTTGTTTCTACTGATCAATG	4380
Db	43321	CCACGACATCGCCCATACACTTCCACAAATAGAGCTAAGTGTGTTTCTACTGATCAATG	4380
OY	4381	CCCCGACGAGTGGATTATTTGTAATGAAAAAAGAAAGCTGGGATTAATCTATACAG	4440
Db	4381	CCCCGACGAGTGGATTATTTGTAATGAAAAAAGAAAGCTGGGATTAATCTATACAG	4440
OY	4441	TGAGTAGACCAATGAGACCAATGTGTGTCACATTACCCTTTTTCCTTTTCTTTTCT	4500
Db	4441	TGAGTAGACCAATGAGACCAATGTGTGTCACATTACCCTTTTTCCTTTTCTTTTCT	4500
OY	4501	TTTTTCTTTTTTTTTTAATGTGAGACAGATCTCATTTCTGTGCTAGGCTGAGTGCAG	4560
Db	4501	TTTTTCTTTTTTTTTTAATGTGAGACAGATCTCATTTCTGTGCTAGGCTGAGTGCAG	4560
OY	4561	TGGGCAATCTCGGCTACTGACAGACCTGCGCCCTGAGGCTCAAGCAATCTCCCAACTC	4620
Db	4561	TGGGCAATCTCGGCTACTGACAGACCTGCGCCCTGAGGCTCAAGCAATCTCCCAACTC	4620
OY	4621	AGCCTCCCAATAGCTGGATCACTGGCAAAACACATGCGCCAGCTAATTTGTATTT	4680
Db	4621	AGCCTCCCAATAGCTGGATCACTGGCAAAACACATGCGCCAGCTAATTTGTATTT	4680
OY	4681	TTTGTAAGACAGAGGTTTCAACATGTGTGCCAGAGCTGAGTCAACCTCTGGGCTCAAGC	4740
Db	4681	TTTGTAAGACAGAGGTTTCAACATGTGTGCCAGAGCTGAGTCAACCTCTGGGCTCAAGC	4740
OY	4741	AATCCTCCTGCTCGGCTCCCAAAAGTGTGGATTACAGATGTAAGCACCGCATCCAG	4800
Db	4741	AATCCTCCTGCTCGGCTCCCAAAAGTGTGGATTACAGATGTAAGCACCGCATCCAG	4800
OY	4801	CCCCACACCTCATTTATACCAATTACCTGCGCAGTACTGTGCACTTTTGCTTCCTAC	4860
Db	4801	CCCCACACCTCATTTATACCAATTACCTGCGCAGTACTGTGCACTTTTGCTTCCTAC	4860
OY	4861	CCCTGCTCTGATCTGGAAGAGAGAGGATTAATGTATAGCTTGTGACAGACAGCCCAATT	4920
Db	4861	CCCTGCTCTGATCTGGAAGAGAGAGGATTAATGTATAGCTTGTGACAGACAGCCCAATT	4920
OY	4921	CAATATTTCTGCGGCAAAAACCTCTTCAAAAAATTAATGTACTTCATTGTATTAATGA	4980
Db	4921	CAATATTTCTGCGGCAAAAACCTCTTCAAAAAATTAATGTACTTCATTGTATTAATGA	4980
OY	4981	ATTACCTTGGAAATGACACCGCCTCAACCTGTTTACATGAGGCTAAATGAAAGGAATTTA	5040
Db	4981	ATTACCTTGGAAATGACACCGCCTCAACCTGTTTACATGAGGCTAAATGAAAGGAATTTA	5040
OY	5041	TAGTCTCTTAATAGGCGTGTACTGACAAAGACCTTGAACACTTTCCAAAGATAGATAT	5100
Db	5041	TAGTCTCTTAATAGGCGTGTACTGACAAAGACCTTGAACACTTTCCAAAGATAGATAT	5100
OY	5101	TTTAAGTCATGCCCCCTTGGCGTGTGCTATGGCACCTTTCCCTCTGAAAGTCGTGTTCTGC	5160
Db	5101	TTTAAGTCATGCCCCCTTGGCGTGTGCTATGGCACCTTTCCCTCTGAAAGTCGTGTTCTGC	5160
OY	5161	CCAATGACCCCTTGGCGTGTGAGCCGAGATGTGACCCCTGCAATAAGGCCCAAGAGAGG	5220
Db	5161	CCAATGACCCCTTGGCGTGTGAGCCGAGATGTGACCCCTGCAATAAGGCCCAAGAGAGG	5220
OY	5221	CTGGGAGTTCCTTCCCTCACTGAAAGAGCCCTTAATTGAATTCACGTGTGAGAGCCCTAGC	5280
Db	5221	CTGGGAGTTCCTTCCCTCACTGAAAGAGCCCTTAATTGAATTCACGTGTGAGAGCCCTAGC	5280
OY	5281	CTCTGCATTTCTGACATTTCCCAACCTCCCAAGCCCTTCACAGACGACTAGTGCCTGC	5340
Db	5281	CTCTGCATTTCTGACATTTCCCAACCTCCCAAGCCCTTCACAGACGAGTATGTTGCTCTGC	5340
OY	5341	ATTTCACACCAAGGTGGAGTATGGCGTTCCTTAAGCGGCTACTGTGCACCACTACCGCAT	5400
Db	5341	ATTTCACACCAAGGTGGAGTATGGCGTTCCTTAAGCGGCTACTGTGCACCACTACCGCAT	5400

QY	5401	CACTGTGGCTCGAAGGACACACAGGGGCATTTTCCCTCACTGAGGGCTCAAAATCCG	5460
QY	5401	CACTGTGGCTCGAAGGACACACAGGGGCATTTTCCCTCACTGAGGGCTCAAAATCCG	5460
Db	5401	CACTGTGGCTCGAAGGACACACAGGGGCATTTTCCCTCACTGAGGGCTCAAAATCCG	5460
QY	5461	TGACAAAGTGGTGGCTCCTGAGACCAAGTATTTCCGAGACGTGGCTCAGTGAAGGGG	5520
Db	5461	TGAGACAAGTGGCTGGCTCCTGAGACCAAGTATTTCCGAGACGTGGCTCAGTGAAGGGG	5520
QY	5521	CCAGCCTGAGGAACCCCTGGCTCTTTTCTTTTAAAGCCAGGCCCCACTTACATAAACAT	5580
Db	5521	CCAGCCTGAGGAACCCCTGGCTCTTTTCTTTTAAAGCCAGGCCCCACTTACATAAACAT	5580
QY	5581	TCAGGCTACTGGAACCACTGAAAGTCCATTTGTTGAAGCTTACTGATGCCAGCCACT	5640
Db	5581	TCAGGCTACTGGAACCACTGAAAGTCCATTTGTTGAAGCTTACTGATGCCAGCCACT	5640
QY	5641	GCTCATCCACGTGGTGGCTGCAAGCCACAGGAAGGCCAGGCATGACAGACTGGGTCT	5700
Db	5641	GCTCATCCACGTGGTGGCTGCAAGCCACAGGAAGGCCAGGCATGACAGACTGGGTCT	5700
QY	5701	AATGCTGTGGTATTGACAGAAAGGAAAGGCTCTCAAGGAAGAGTCACTGGGACAAACA	5760
Db	5701	AATGCTGTGGTATTGACAGAAAGGAAAGGCTCTCAAGGAAGAGTCACTGGGACAAACA	5760
QY	5761	CAAGCCCAACCGACATGGCTTGGTGAAGGTTAGACAGACTGCTGTGTGTGATCTGCAGT	5820
Db	5761	CAAGCCCAACCGACATGGCTTGGTGAAGGTTAGACAGACTGCTGTGTGTGATCTGCAGT	5820
QY	5821	GCTTCACAGGAATATTTATTTATTCATGCGAGTACGTTTAAAGTGGCATTTTATTCATTTT	5880
Db	5821	GCTTCACAGGAATATTTATTTATTCATGCGAGTACGTTTAAAGTGGCATTTTATTCATTTT	5880
QY	5881	CTGTCTTTTAAATAACAATATGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTGG	5940
Db	5881	CTGTCTTTTAAATAACAATATGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTGG	5940
QY	5941	CTACTTGTCCCTGGTTCAGTAGAGGCCCGGTTTCCAGTTGTTGACTGTGACAGGCTC	6000
Db	5941	CTACTTGTCCCTGGTTCAGTAGAGGCCCGGTTTCCAGTTGTTGACTGTGACAGGCTC	6000
QY	6001	AGCATGGGCTCGACAGATGCTCTTAATTTTGTGATGATACAGAAAGCCAGGCTTGGG	6060
Db	6001	AGCATGGGCTCGACAGATGCTCTTAATTTTGTGATGATACAGAAAGCCAGGCTTGGG	6060
QY	6061	ATACAAAGTCTTCTTCCCTCTTCATTTTATGACCCTGGACACTGTGTGAACAGATGTTTGTCC	6120
Db	6061	ATACAAAGTCTTCTTCTTCCCTCTTCATTTTATGACCCTGGACACTGTGTGAACAGATGTTTGTCC	6120
QY	6121	GGAATTAATAATATATGCTTGGAGTCTGCCCAAAAAAAAAAAAA 6163	
Db	6121	GGAATTAATAATATATGCTTGGAGTCTGCCCAAAAAAAAAAAAA 6163	
RESULT 2			
US-09-899-569A-1			
; Sequence 1, Application US/09899569A			
; Patent No. US20020142003A1			
GENERAL INFORMATION:			
; APPLICANT: NO. US20020142003Albert Schweitzer			
; APPLICANT: Marwa Scherl-Moslager			
; APPLICANT: Wolfgang Sommerhuber			
; TITLE OF INVENTION: Tumorssoziliertes Antigen (3345)			
; FILE REFERENCE: 0652.2280001			
; CURRENT APPLICATION NUMBER: US/09/899,569A			
; PRIOR APPLICATION NUMBER: DE 100 33 080.0			
; PRIOR FILING DATE: 2000-07-07			
; PRIOR APPLICATION NUMBER: DE 101 19 294.0			
; PRIOR FILING DATE: 2001-04-19			
; PRIOR APPLICATION NUMBER: US 60/243,158			
; PRIOR FILING DATE: 2000-10-25			
; PRIOR APPLICATION NUMBER: US 60/297,747			

;; PRIOR FILING DATE: 2001-06-14
;; NUMBER OF SEQ ID NOS: 40
;; SOFTWARE: Patentin Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 5897
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (5348)..(5348)
;; OTHER INFORMATION: n is any nucleotide of a, t, g or c
;; NAME/KEY: misc_feature
;; LOCATION: (5360)..(5360)
;; OTHER INFORMATION: n is any nucleotide of a, t, g or c
;; NAME/KEY: misc_feature
;; LOCATION: (5425)..(5425)
;; OTHER INFORMATION: n is any nucleotide of a, t, g or c
;; NAME/KEY: misc_feature
;; LOCATION: (5438)..(5438)
;; OTHER INFORMATION: n is any nucleotide of a, t, g or c
;; NAME/KEY: 5'UTR
;; LOCATION: (1)..(214)
;; NAME/KEY: CDS
;; LOCATION: (215)..(2464)
;; NAME/KEY: 3'UTR
;; LOCATION: (2465)..(5897)
US-09-899-569A-1

Query Match 95.0%; Score 5854.2; DB 10; Length 5897;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5865; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 291 CCTGAACCTGGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGGTGGCGGGCGCT 350
DB 25 CCTGAACCTGGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGGTGGCGGGCGCT 84
OY 351 GCGGGCGGGGAGAAAGCTTTGAGATTGCTGTCGACAGAGAAAGCAATTAAGTTCT 410
DB 85 GCGGGCGGGGAGAAAGCTTTGAGATTGCTGTCGACAGAGAAAGCAATTAAGTTCT 144
OY 411 CATTAAGCTGGGGACCCGACCTCTGCTGGCAAAACCCCTTTACATGTCATTTCTAAAG 470
DB 145 CATTAAGCTGGGGACCCGACCTCTGCTGGCAAAACCCCTTTACATGTCATTTCTAAAG 204
OY 471 AATATTAACCATGTTGTCCATCAAGCTGAGAGAAAGATAGTCTTACCTTTAGTGCCA 530
DB 205 AATATTAACCATGTTGTCCATCAAGCTGAGAGAAAGATAGTCTTACCTTTAGTGCCA 264
OY 531 GAGTCTGGAATCACTTTGTCTATAGATGATCCAGAAATATTTGACTGTATGTCAGGGCC 590
DB 265 GAGTCTGGAATCACTTTGTCTATAGATGATCCAGAAATATTTGACTGTATGTCAGGGCC 324
OY 591 ATGTCCTTTTGGGAGGTTGACCTTCAGCCCTCGACATCGTTGTTGCCCTCAACAG 650
DB 325 ATGTCCTTTTGGGAGGTTGACCTTCAGCCCTCGACATCGTTGTTGCCCTCAACAG 384
OY 651 AACTTTCACTGTGGATGTCAAAGCTCATATAGAGCATCGGTTTAGAGCTGACGTTTTCAT 710
DB 385 AACTTTCACTGTGGATGTCAAAGCTCATATAGAGCATCGGTTTAGAGCTGACGTTTTCAT 444
OY 711 CCTTCGCGGAGAGATGATGATCGGAGTGAAGCTGAGAGCGGAGTCACTGATCCAT 770
DB 445 CCTTCGCGGAGAGATGATGATCGGAGTGAAGCTGAGAGCGGAGTCACTGATCCAT 504
OY 771 CAGCGGCGGAATCGATGCGACCGTGTGATCGAATCGAATCTTCTGAGCAATGGCACTGT 830
DB 505 CAGCGGCGGAATCGATGCGACCGTGTGATCGAATCGAATCTTCTGAGCAATGGCACTGT 564
OY 831 GTCCCGGATCAAGATGCAAGAGAGAGTGAATAATGGCTTACACCTCCATGGTTCCACCC 890
DB 565 GTCCCGGATCAAGATGCAAGAGAGAGTGAATAATGGCTTACACCTCCATGGTTCCACCC 624
OY 891 CAGAAATGTCCTCGGCTTCAGCATTTGCAAAACGCTCATGTATAAAACGCTGTGCATCAT 950

DB 625 CAGAAATGTCCTCGGCTTCAGCATTTGCAAAACGCTCATGTATAAAACGCTGTGCATCAT 684
OY 951 CGAGTCTGTGTTGAGGGTGAAGGCTCAGCAACCTGATGTGTGCCACTACCCAGAGG 1010
DB 685 CGAGTCTGTGTTGAGGGTGAAGGCTCAGCAACCTGATGTGTGCCACTACCCAGAGG 744
OY 1011 CTTCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1070
DB 745 CTTCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804
OY 1071 CGTCTCTCTCTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 1130
DB 805 CGTCTCTCTCTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 864
OY 1131 CTACATCCGCGGCTCCAGCAACCCGAGGTGTTCAAGCTGAGAGCAAGCCTGG 1190
DB 865 CTACATCCGCGGCTCCAGCAACCCGAGGTGTTCAAGCTGAGAGCAAGCCTGG 924
OY 1191 GAACATGGGGGAGACTTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAGTCC 1250
DB 925 GAACATGGGGGAGACTTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAGTCC 984
OY 1251 AGGATCTCTCGGCTGCAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGA 1310
DB 985 AGGATCTCTCGGCTGCAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGA 1044
OY 1311 AATCTAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1370
DB 1045 AATCTAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104
OY 1371 CGTCAACAGAGCCGCAAGTTGTCTCCCTGGCTGTTCGTGTGTGTGTGTGTGTGTGTGTGT 1430
DB 1105 CGTCAACAGAGCCGCAAGTTGTCTCCCTGGCTGTTCGTGTGTGTGTGTGTGTGTGTGTGT 1164
OY 1431 CAGTAGCACTCACCTGACATCTGCTGCTCAAAACCAAAATCTCTCTTGTGATGA 1490
DB 1165 CAGTAGCACTCACCTGACATCTGCTGCTCAAAACCAAAATCTCTCTTGTGATGA 1224
OY 1491 TCTGACAGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1550
DB 1225 TCTGACAGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1284
OY 1551 CCAAGGAATCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1610
DB 1285 CCAAGGAATCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1344
OY 1611 GCATGACTCTCTCTGGAAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670
DB 1345 GCATGACTCTCTCTGGAAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1404
OY 1671 AGCCGAGAGCTGACAGCATACACAGAGAACCCCTGCAACACAGCTTACGTAAGT 1730
DB 1405 AGCCGAGAGCTGACAGCATACACAGAGAACCCCTGCAACACAGCTTACGTAAGT 1464
OY 1731 CGTGGCAATGCTATACCCAGAGAGCTGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1790
DB 1465 CGTGGCAATGCTATACCCAGAGAGAGCTGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1524
OY 1791 TATCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1850
DB 1525 TATCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1584
OY 1851 CTTCGAAGAAGAGCTCCAGGAGAGGTGTGAAGGTGTCTTATTAATCTTATTTCAAGA 1910
DB 1585 CTTCGAAGAAGAGCTCCAGGAGAGGTGTGAAGGTGTCTTATTAATCTTATTTCAAGA 1644
OY 1911 GGAAGGCTTTTACAGGCTGACCCCTGACACAAAGAGGTCTACTAGAGACCCCA 1970
DB 1645 GGAAGGCTTTTACAGGCTGACCCCTGACACAAAGAGGTCTACTAGAGACCCCA 1704
OY 1971 CTGGGACCGGGGCTGCACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030

Db 1705 CTGGAGCCGGGGCTGCCATCCCTCACCTGTGTCTGGACATCAGCGTGCCAGAGA 1764
QY 2031 CCAAGTGGCCCTGCTGACTTCTTTTAAAGAGCGGAGCGGGGTGTCTGCCAGAGCGGC 2090
Db 1765 CCAAGTGGCCCTGCTGACTTCTTTTAAAGAGCGGAGCGGGGTGTCTGCCAGAGCGGC 1824
QY 2091 CGCATTCATGATCATCCAGAGAGCAGCGGACCCGGGCTGAGAGATCTTTCAGCCTGAGCA 2150
Db 1825 CGCATTCATGATCATCCAGAGAGCAGCGGACCCGGGCTGAGAGATCTTTCAGCCTGAGCA 1884
QY 2151 GGATGTGCTCCCCAAGCCAGCTTCCACATCCAGCTTCGGGTCAACATCTCTACTG 2210
Db 1885 GGATGTGCTCCCCAAGCCAGCTTCCACATCCAGCTTCCGGGTCAACATCTCTACTG 1944
QY 2211 CAGCCCCAGAGCGGCGAAGCAGCTAGACCTGTCTTCTCGGTGACACTTACCCCAAGAC 2270
Db 1945 CAGCCCCAGAGCGGCGAAGCAGCTAGACCTGTCTTCTCGGTGACACTTACCCCAAGAC 2004
QY 2271 TGTGGACTTGACTGTCACTCTCATCGACGGGTGGAGGTGAGTCTTACTGCTGTGC 2330
Db 2005 TGTGGACTTGACTGTCACTCTCATCGACGGGTGGAGGTGAGTCTTACTGCTGTGC 2064
QY 2331 CCTGGGCTCATCTTGTGTGTGAAAAAAGAAAAAAGAAAAAGCAAGGCGCCGCG 2390
Db 2065 CCTGGGCTCATCTTGTGTGTGAAAAAAGAAAAAAGAAAAAGCAAGGCGCCGCG 2124
QY 2391 TGTGGGTATCTACATGGAACAATCACTAGATGCGAGCGCAGCCAAAAAGTTTCA 2450
Db 2125 TGTGGGTATCTACATGGAACAATCACTAGATGCGC - AGGAGCGCAAAAAAGTTTCA 2183
QY 2451 GAAAAGCGGAAAGACATGACTCCCATGTGTATGACGTATCGAGACACCATGTATTA 2510
Db 2184 GAAAAGCGGAAAGACATGACTCCCATGTGTATGACGTATCGAGACACCATGTATTA 2243
QY 2511 TGGGCACTGTGCTCAGGATTCAGGCGGCTCCTTCGACAGCAGAGGTGAGACCTTCCG 2570
Db 2244 TGGGCACTGTGCTCAGGATTCAGGCGGCTCCTTCGACAGCAGAGGTGAGACCTTCCG 2303
QY 2571 GCCGTTCAGAGGCGACATGAGGGGTCTGTCCCTCCCAACCCACATATGCTCCAGGC 2630
Db 2304 GCCGTTCAGAGGCGACATGAGGGGTCTGTCCCTCCCAACCCACATATGCTCCAGGC 2363
QY 2631 CCCACCTGCAAAAGTTGGCCACTGAGAGCCACCTCCCTCCCTAAGTCTGAGAG 2690
Db 2364 CCCACCTGCAAAAGTTGGCCACTGAGAGCCACCTCCCTCCCTAAGTCTGAGAG 2423
QY 2691 TGAACGCTACACCTTCCCATCCCAACAATGGGGATGTAAGAGCAAGAGACAGCAT 2750
Db 2424 TGAACGCTACACCTTCCCATCCCAACAATGGGGATGTAAGAGCAAGAGACAGCAT 2483
QY 2751 TCCCTTACTGACACTGAGAGCCCATGAGCCAGCAAAATACTTGATTCATTCAGAC 2810
Db 2484 TCCCTTACTGACACTGAGAGCCCATGAGCCAGCAAAATACTTGATTCATTCAGAC 2543
QY 2811 GCTTGTGAGTTTCATAAAGCAGGGCAGTGAACACCCGTCGGTCTTCTAACAGAAA 2870
Db 2544 GCTTGTGAGTTTCATAAAGCAGGGCAGTGAACACCCGTCGGTCTTCTAACAGAAA 2603
QY 2871 TCCCTAAAGAGAGAAATTATACAGAAAGAACAGAGAGTTTTCGAGGACCGCCAAC 2930
Db 2604 TCCCTAAAGAGAGAAATTATACAGAAAGAACAGAGAGTTTTCGAGGACCGCCAAC 2663
QY 2931 TTCACATTTGCTCAGTGAATCTCTAAGGGCAAGACATTAATAATGATTCATC 2990
Db 2664 TTCACATTTGCTCAGTGAATCTCTAAGGGCAAGACATTAATAATGATTCATC 2723
QY 2991 TGGATACAGTCAATGACAGCTCATGTGCTCTCAACTTAAAGCTGTGCGTTAGCCAGCTG 3050
Db 2724 TGGATACAGTCAATGACAGCTCATGTGCTCTCAACTTAAAGCTGTGCGTTAGCCAGCTG 2783
QY 3051 TAAATGAGAGAGAGGCGCTGAGTCACTTACATTAAGGTTGACAGCAAGCCCTGGATTGAG 3110
Db 2784 TAAATGAGAGAGAGGCGCTGAGTCACTTACATTAAGGTTGACAGCAAGCCCTGGATTGAG 2843

QY 3111 AGTGTAAACAGAGGCTTGGCCCTCTTCAGACAAACATTCCAATTCGAAGAGCTTACCT 3170
Db 2844 AGTGTAAACAGAGGCTTGGCCCTCTTCAGACAAACATTCCAATTCGAAGAGCTTACCT 2903
QY 3171 GAGTCCCTACTCTCACTGGGGTCCCGAGATGAAAAAGCAATGTGCTTTTATATAT 3230
Db 2904 GAGTCCCTACTCTCACTGGGGTCCCGAGATGAAAAAGCAATGTGCTTTTATATAT 2963
QY 3231 ATTTATTTGGTGGTCCGTGTATTTTAAAGATCAATATATAACACCTAGCTTTTC 3290
Db 2964 ATTTATTTGGTGGTCCGTGTATTTTAAAGATCAATATATAACACCTAGCTTTTC 3023
QY 3291 ACCTGACTAGTATAACTACTACTAATCTGTTGATGCTGGGTTGTACTTCTACTG 3350
Db 3024 ACCTGACTAGTATAACTACTACTAATCTGTTGATGCTGGGTTGTACTTCTACTG 3083
QY 3351 ACCGCTAGATTAACGTTGGCTGTCCCGCAGGTGGTGGGAATATTTACAATCTGTCCA 3410
Db 3084 ACCGCTAGATTAACGTTGGCTGTCCCGCAGGTGGTGGGAATATTTACAATCTGTCCA 3143
QY 3411 CCAGAAAAAGATGTGTGTTTGAAGCATATGTCCTTGTGATTAAGACTT 3470
Db 3144 CCAGAAAAAGATGTGTGTTTGAAGCATATGTCCTTGTGATTAAGACTT 3203
QY 3471 CCGTATTCCTAGGTGGTTCGTGTATCCATTTGTGAAATTCATCTTGAATCCATT 3530
Db 3204 CCGTATTCCTAGGTGGTTCGTGTATCCATTTGTGAAATTCATCTTGAATCCATT 3263
QY 3531 GTCTATAGTCTTACGCAATTAAGAAATTTTCACATGTCGCTTCCCTAGC 3590
Db 3264 GTCTATAGTCTTACGCAATTAAGAAATTTTCACATGTCGCTTCCCTAGC 3323
QY 3591 TGCAGCAATCTTGTGACATTTAAAGAGAAATTAAGAAATATCTCATCTTAAATG 3650
Db 3324 TGCAGCAATCTTGTGACATTTAAAGAGAAATTTAAAGAAATATCTCATCTTAAATG 3383
QY 3651 TTTAAATATATACCAACAGTGGCCCTCATTTAGTTTCTGTGACACTGCAACCAT 3710
Db 3384 TTTAAATATATACCAACAGTGGCCCTCATTTAGTTTCTGTGACACTGCAACCAT 3443
QY 3711 TACTTGTACTTAAAAACAACATTAAGTTATAGTCTGGGGATTAAGAAATTCCTAAAT 3770
Db 3444 TACTTGTACTTAAAAACAACATTAAGTTATAGTCTGGGGATTAAGAAATTCCTAAAT 3503
QY 3771 GGATGTCCCTGAATGAATAACAGGTGTGACAGAGCTGTGCTCTCTGAAGGCTTAG 3830
Db 3504 GGATGTCCCTGAATGAATAACAGGTGTGACAGAGCTGTGCTCTCTGAAGGCTTAG 3563
QY 3831 GGAGAAAGCCGTTCTTGGCATTTCAAGCTTTAGAGGCTGGCTGATTCACAGGCTCA 3890
Db 3564 GGAGAAAGCCGTTCTTGGCATTTCAAGCTTTAGAGGCTGGCTGATTCACAGGCTCA 3623
QY 3891 GTGGCTGTGTAAGCTTTTTCACATGGCATGCTGTGACACTGTGGCCCTCCACTTCCCTC 3950
Db 3624 GTGGCTGTGTAAGCTTTTTCACATGGCATGCTGTGACACTGTGGCCCTCCACTTCCCTC 3683
QY 3951 TTTGACTTACAAAGCCACAGAGAAATCCAGGATTAATCTCTCATATTAAGATCTTCA 4010
Db 3684 TTTGACTTACAAAGCCACAGAGAAATCCAGGATTAATCTCTCATATTAAGATCTTCA 3743
QY 4011 TCATCTGGAAGAGCCTTTTGGCATGCAACAACATAGCCAGAGTGGGATTAAGACC 4070
Db 3744 TCATCTGGAAGAGCCTTTTGGCATGCAACAACATAGCCAGAGTGGGATTAAGACC 3803
QY 4071 AGGACATCTTTGGGGTGTCTTATTTGCTTACACACCTTCTGCACTGACTCCACA 4130
Db 3804 AGGACATCTTTGGGGTGTCTTATTTGCTTACACACCTTCTGCACTGACTCCACA 3863
QY 4131 GGAGAGGCTCAAAATGATGTGGGCGACAGGAGTGTGTTTGTAGCTTGGGACTGTAAAC 4190
Db 3864 GGAGAGGCTCAAAATGATGTGGGCGACAGGAGTGTGTTTGTAGCTTGGGACTGTAAAC 3923

4191 ACTTAAAAAACCCAGATCAGAAATGCTGGCCATGCTGGGCTCACTTTCACCTAGC 4250
4251 AACAACTGGCTGGAGCTGGGACACAGCTCTGCTTTAGAAGGGGTGCCACTTTCACAG 4310
3984 AACAACTGGCTGGAGCTGGGACACAGCTCTGCTTTAGAAGGGGTGCCACTTTCACAG 4043
4311 TCACCAACAGCCACACTACGCCCTATCACTTCCCAATAGAGCTTAAGTGTGTTCTTA 4370
4044 TCACCAACAGCCACACTACGCCCTATCACTTCCCAATAGAGCTTAAGTGTGTTCTTA 4103
4371 CTGATCAATGCCCCCTGAGTGGCTTATTTGTAATGAAAAAGAGCTGGGATTAATC 4430
4104 CTGATCAATGCCCCCTGAGTGGCTTATTTGTAATGAAAAAGAGCTGGGATTAATC 4163
4431 TCTAATCAGGTAGTAGACATAGACCAATGTGTCTCACTTACCTTTTCTTTTCTT 4490
4164 TCTAATCAGGTAGTAGACATAGACCAATGTGTCTCACTTACCTTTTCTTTTCTT 4223
4491 TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTT 4550
4224 TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTT 4283
4551 TGGAGTGCAGTGGCGCAATCTGGGCTCACTGCAACCTGCTGCTGGGCTCAAGCAAT 4610
4284 TGGAGTGCAGTGGCGCAATCTGGGCTCACTGCAACCTGCTGCTGGGCTCAAGCAAT 4343
4611 CTCACCACTCAGGCTCCCAATAGCTGGGATCACTGGCACAACAACCACTGCCCAGCTTA 4670
4344 CTCACCACTCAGGCTCCCAATAGCTGGGATCACTGGCACAACAACCACTGCCCAGCTTA 4403
4671 TTTTGTATTTTGTATGAGACAGAGGTTTCAACATGTTGCCAGGCTGCTCACTCTCT 4730
4404 TTTTGTATTTTGTATGAGACAGAGGTTTCAACATGTTGCCAGGCTGCTCACTCTCT 4463
4731 GGGCTCAAGCAATCTCTGCTGCGCTGCGCTGCCAAGAGTGGGATTAAGAGATGAGCCA 4790
4464 GGGCTCAAGCAATCTCTGCTGCGCTGCGCTGCCAAGAGTGGGATTAAGAGATGAGCCA 4523
4791 CCGCATCAGCCCGCACACCTCATTTTATACCAATTAACCTGCCCCAGTAACTGTGGACTTTT 4850
4524 CCGCATCAGCCCGCACACCTCATTTTATACCAATTAACCTGCCCCAGTAACTGTGGACTTTT 4583
4851 GCTTCTCTACCCCTGCTGCTGATCTGGAAGAGAGGATTAATGTTAATAGCTTGTACACCA 4910
4584 GCTTCTCTACCCCTGCTGCTGATCTGGAAGAGAGGATTAATGTTAATAGCTTGTACACCA 4643
4911 GTCCCAAGTTCATATTTTGTGGGGCAAAAACCTTCTCAAAAAATTAATGTACTGATG 4970
4644 GTCCCAAGTTCATATTTTGTGGGGCAAAAACCTTCTCAAAAAATTAATGTACTGATG 4703
4971 TATTCATGAATTAACCTTGAATGACACCGCTCAACTTGTTCACATGGCATAAATGAA 5030
4704 TATTCATGAATTAACCTTGAATGACACCGCTCAACTTGTTCACATGGCATAAATGAA 4763
5031 AGGAATTTTATAGTCTGCTTAATGAGGCTGACTGCAAGACCTTGTGAACACTTTCAGAG 5090
4764 AGGAATTTTATAGTCTGCTTAATGAGGCTGACTGCAAGACCTTGTGAACACTTTCAGAG 4823
5091 GATAGGATTTTATAGTCTGAGCCCTTGGGCTGTGCTATGGACACTTCCCTTCGAAGTC 5150
4824 GATAGGATTTTATAGTCTAGCCCTTGGGCTGTGCTATGGACACTTCCCTTCGAAGTC 4883
5151 TGGTTCCTCCAGAGTACCCCTTGGGCTTGTGAGCCGAGATGCTGACCTGTCATAAAGGC 5210
4884 TGGTTCCTCCAGAGTACCCCTTGGGCTTGTGAGCCGAGATGCTGACCTGTCATAAAGGC 4943
5211 CAAAGAGGCTGCGGCTTCTCTCTCTCACTGACAGAGCCCTTATTTGAATTCACCTGTG 5270
4944 CAAAGAGGCTGCGGCTTCTCTCTCTCACTGACAGAGCCCTTATTTGAATTCACCTGTG 5003
5271 GAGCCCTAGCCCTCATTTCTCGACATTTCCCAACCTCCAGCCCTTCCAGAGAGACTA 5330

5004 GAGCCCTAGCCCTCATTTCTCGACATTTCCCAACCTCCAGCCCTTCCAGAGAGACTA 5063
5331 GGTGCCCTGCATTCACCCCAAGGTGGGATTTGCTTCTTAGGCTGCTACTTGTACCA 5390
5064 GGTGCCCTGCATTCACCCCAAGGTGGGATTTGCTTCTTAGGCTGCTACTTGTACCA 5123
5391 TCACCGACATCACTGTTGCTGCAAGGACACAGGTGGCATTTTCTTCACTGAGGGC 5450
5124 TCACCGACATCACTGTTGCTGCAAGGACACAGGTGGCATTTTCTTCACTGAGGGC 5183
5451 TCAAAACCTCGGACAAAGTTGCTGGCTCCGAGAACAGTATTTCTGAGAGCTGGCTCA 5510
5184 TCAAAACCTCGGACAAAGTTGCTGGCTCCGAGAACAGTATTTCTGAGAGTGGCTCA 5243
5511 GTGAAGGGGCCAGCCCTGAGGAACCTGCTCTTTCTTTTAAAGCCAGGCCCACTTAC 5570
5244 GTGAAGGGGCCAGCCCTGAGGAACCTGCTCTTTCTTTTAAAGCCAGGCCCACTTAC 5303
5571 ATTAACATTTTCAGGCTCACTGGAACAGTGAAGTGGCATTTGTGAAGCTTACTGATG 5630
5304 ATTAACATTTTCAGGCTCACTGGAACAGTGAAGTGGCATTTGTGAAGCTTACTGATG 5363
5631 CCAGCCACTGCTCATTCACAGTGGTGGCTGCTGCTATGCTACGAGGAAGCCAGGCTAGG 5690
5364 CCAGCCACTGCTCATTCACAGTGGTGGCTGCTGCTATGCTATCCAGAGAGGCCAGGCTAGG 5423
5691 ACTGCTCTTAATGCTGTGTGCTATTGTCACAGGAAGGAAAGTCTCAAGGAAGTCACT 5750
5424 ANTGGTCTTAATGNTGTGTGCTATTGTCACAGGAAGGAAAGTCTCAAGGAAGTCACT 5483
5751 GGGACAAAGCAAGCCCGGACATGGCTTGGTAAAGTTAGCAGACTGGTGTGTG 5810
5484 GGGACAAAGCAAGCCCGGACATGGCTTGGTAAAGTTAGCAGACTGGTGTGTG 5543
5811 GATCTGAGTGTCTCACTGGAATTAATTTATTCATTCAGATACTTTTATAGTGGCATTT 5870
5544 GATCTGAGTGTCTCACTGGAATTAATTTATTCATTCAGATACTTTTATAGTGGCATTT 5603
5871 TATTCATTTCTGCTTAAATTAACAAATGTAACAAATTAATTAATTAATTAATTAATTA 5930
5604 TATTCATTTCTGCTTAAATTAACAAATGTAACAAATTAATTAATTAATTAATTAATTA 5663
5931 AGTGTCTGCTACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5990
5664 AGTGTCTGCTACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5723
5991 TGACAGGCTCAGATGGGCTCAGCAGATGCTGCTTAATTTGAGATGATACGAAAGCC 6050
5724 TGACAGGCTCAGATGGGCTCAGCAGATGCTGCTTAATTTGAGATGATACGAAAGCC 5783
6051 AGGCTTTGGGATCAAGTCTTCTCTCTCACTTGAATGCGGTGCTGCTGCTGCTGCTGCTG 6110
5784 AGGCTTTGGGATCAAGTCTTCTCTCTCACTTGAATGCGGTGCTGCTGCTGCTGCTGCTG 5843
6111 GTTTTGTCCGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6163
5844 GTTTTGTCCGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5896

RESULT 3
US-09-922-217-362
; Sequence 362, Application US/09922217
; Patent No. US20020076141A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaochun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongfong
; APPLICANT: Jiang, Yuxin

APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aljun
APPLICANT: Clapper, Jonathan D.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C13
CURRENT APPLICATION NUMBER: US/09/922.217
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 362
LENGTH: 544
TYPE: DNA
ORGANISM: Homo sapiens
US-09-922-217-362

Query Match 8.8%; Score 542.4; DB 10; Length 544;
Best Local Similarity 99.8%; Pred. No. 2.1e-102;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3068 CCGAGTCACCTAGACATAGAGTTGACAGACCCCTGGATTACAGAGTTAAACAGAGGCT 3127
1 CCGAGTCACCTAGACATAGAGTTGACAGACCCCTGGATTACAGAGTTAAACAGAGGCT 60
-3128 TGGCCCTTCAGAGCAACAGTCCCAATTCAGAGAGCTACAGGTCCTACTCTCAC 3187
61 TGGCCCTTCAGAGCAACAGTCCCAATTCAGAGAGCTACAGGTCCTACTCTCAC 120
3188 TGGGGTCCCGAGATGAAAAAGACATATGCTTTTATTTATTTATTTATTTGGTGCTCT 3247
121 TGGGGTCCCGAGATGAAAAAGACATATGCTTTTATTTATTTATTTATTTGGTGCTCT 180
3248 GTGTATTATTAAGATCAATGTATTAACACCTAGCTCTTTTACCTGACTTAGTAATAA 3307
181 GTGTATTATTAAGATCAATGTATTAACACCTAGCTCTTTTACCTGACTTAGTAATAA 240
3308 CTCATACTAATGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3367
241 CTCATACTAATGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
3368 TGGCTGTCCTCCAGAGTGGTGGGAATTAATTAACAATCTGTCACCAACCAAAAAGATGTG 3427
301 TGGCTGTCCTCCAGAGTGGTGGGAATTAATTAACAATCTGTCACCAACCAAAAAGATGTG 360
3428 TGTGTGAGCAGCATGTGACATATCTGCTTTGATTAAGAGACTTCCTGATTTCTAGGTG 3487
361 TGTGTGAGCAGCATGTGACATATCTGCTTTGATTAAGAGACTTCCTGATTTCTAGGTG 420
3488 GTTCGTGTTATCCCATTTGGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3547
421 GTTCGTGTTATCCCATTTGGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 480
3548 ATAAAGAAATTTCCCAATTCATCTTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3607
481 ATAAAGAAATTTCCCAATTCATCTTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 540
3608 ATTTT 3611
541 ATTT 544

RESULT 4
US-09-833-263-362
Sequence 362, Application US/09833263
Patent No. US2002010547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aljun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeline J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE

FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833.263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 362
LENGTH: 544
TYPE: DNA
ORGANISM: Homo sapien
US-09-833-263-362

Query Match 8.8%; Score 542.4; DB 10; Length 544;
Best Local Similarity 99.8%; Pred. No. 2.1e-102;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3068 CCGAGTCACCTAGACATAGAGTTGACAGACCCCTGGATTACAGAGTTAAACAGAGGCT 3127
1 CCGAGTCACCTAGACATAGAGTTGACAGACCCCTGGATTACAGAGTTAAACAGAGGCT 60
3128 TGGCCCTTCAGAGCAACAGTCCCAATTCAGAGAGCTACAGGTCCTACTCTCAC 3187
61 TGGCCCTTCAGAGCAACAGTCCCAATTCAGAGAGCTACAGGTCCTACTCTCAC 120
3188 TGGGGTCCCGAGATGAAAAAGACATATGCTTTTATTTATTTATTTATTTGGTGCTCT 3247
121 TGGGGTCCCGAGATGAAAAAGACATATGCTTTTATTTATTTATTTATTTGGTGCTCT 180
3248 GTGTATTATTAAGATCAATGTATTAACACCTAGCTCTTTTACCTGACTTAGTAATAA 3307
181 GTGTATTATTAAGATCAATGTATTAACACCTAGCTCTTTTACCTGACTTAGTAATAA 240
3308 CTCATACTAATGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3367
241 CTCATACTAATGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
3368 TGGCTGTCCTCCAGAGTGGTGGGAATTAATTAACAATCTGTCACCAACCAAAAAGATGTG 3427
301 TGGCTGTCCTCCAGAGTGGTGGGAATTAATTAACAATCTGTCACCAACCAAAAAGATGTG 360
3428 TGTGTGAGCAGCATGTGACATATCTGCTTTGATTAAGAGACTTCCTGATTTCTAGGTG 3487
361 TGTGTGAGCAGCATGTGACATATCTGCTTTGATTAAGAGACTTCCTGATTTCTAGGTG 420
3488 GTTCGTGTTATCCCATTTGGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3547
421 GTTCGTGTTATCCCATTTGGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 480
3548 ATAAAGAAATTTCCCAATTCATCTTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 3607
481 ATAAAGAAATTTCCCAATTCATCTTGAATTCATCTTGAATCCCATTTGCTCTATAGTCTAGCA 540
3608 ATTTT 3611
541 ATTT 544

RESULT 5
US-09-815-343-1335/C
Sequence 1335, Application US/09815343
Patent No. US20010055596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeline
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815.343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1335
LENGTH: 541


```
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(636)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-370
```

```
Query Match
Best Local Similarity 90.3%; Pred. No. 1.9e-91; Length 636;
Matches 577; Conservative 0; Mismatches 54; Indels 8; Gaps 6;
```

```
QY 5270 GGAACCTTACGCTTCATTCGACATTCGCCAACCTCCAGCCCTTCCAGAGAGCT 5329
Db 636 GGAAGCCNANNCCTCAATTTAAATATCCCAACCTT--CAAGCCCTTCCAAAGAGANT 579
QY 5330 AGGAGCCCTGATTCACCCCAAGGTGGGATTTGGCTTCCCTT--AGGCTGGCTACTTG-TC 5386
Db 578 A-GAGCCCTGANTCCNCCCAAGNAGGATTTGGCTTCCCTTAAAGNNGTINACTTGTGTC 520
QY 5387 ACCATCAACCGACATCACTGTTCCTGCAAGGACACACAGCTGGCCATTTTCTTC-AACTG 5445
Db 519 ACNATCAAGGAATNNCTGGTGTTCGANGSACNCCAAAGGGCCATTTTCTTCAANTTA 460
QY 5446 AGGCTCAAAACCTCTGGACAAAGTTGCTGCTCC-TGAGACCAATTTTCTTGAGCTGT 5504
Db 459 AGGCTCAAAACCTCTGGACAAAGTTGCTGCTCCCTTAAGNCCAGATTTTCTGAAAGCTG 400
QY 5505 GCCCTCAAGAAAGGGGCCAGGCTGAGAACCTGGCTCTTTCTTTAAAGCCAGGCCCC 5554
Db 399 CCTAAGAAAGGGGCCAGGCTTGAAGAACCTGGCTCTTTCTTTAAAGCCAGGCCCC 340
QY 5565 ACTTACATTAATAATTTTCAAGGCTCACTGGAACAGTGAAGTGCATTTGTGTAAGCTTAC 5624
Db 339 ACTTACGTAATAATTTTCAAGGCTCACTGGAACAGTGAAGTGCATTTGTGTAAGCTTAC 280
QY 5625 TGCATGCGACGCCCACTGCTCTATCCAGTGTGTGCTGCTGCTTACGAGAGAGGCCGCA 5684
Db 279 TGCATGCGACGCCCACTGCTCTATCCAGTGTGTGCTGCTGCTTACGAGAGAGGCCGCA 220
QY 5685 TGCAGCATGCTCTATGCTGTGTGCTATTCGACAGAAAGGAAAGGCTCAAGGAAGAG 5744
Db 219 TGCAGCATGCTCTATGCTGTGTGCTATTCGACAGAAAGGAAAGGCTCAAGGAAGAG 160
QY 5745 TCAACTGGGACAGACAGACAGCCACCGACATGGCCTTGTAAAGGTTAGCAGACTGGTG 5804
Db 159 TCAACTGGGACAGACAGACAGCCACCGACATGGCCTTGTAAAGGTTAGCAGACTGGTG 100
QY 5805 TGTGTGATCTGCAAGTGTCTTACATGGAATTAATTTATTCATTCAGATACTTTTAGGTG 5864
Db 99 TGTGTGATCTGCAAGTGTCTTACATGGAATTAATTTATTCATTCAGATACTTTTAGGTG 40
QY 5865 GCATTTTATTCATTTCCGTGCTTTAATAATAAATAATGT 5903
Db 39 GCATTTTATTCATTTCTCTGTGCTTTAATAATAAATAATGT 1
```

```
RESULT 8
US-09-815-343-393
; Sequence 393, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815.343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 393
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(535)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-393
```

```
Query Match
Best Local Similarity 94.8%; Pred. No. 5.2e-90; Length 535;
Matches 509; Conservative 0; Mismatches 26; Indels 2; Gaps 1;
```

```
QY 3068 CCTAGTCACCTACATAGAGGTTCAGACAGCCCTGATTCAGAGTGTAAACAGAGCT 3127
Db 1 CCTAGTCACCTACATAGAGGTTCAGACAGCCCTGATTCAGAGTGTAAACAGAGCT 60
QY 3128 TGCCTCTTCAAGCAACAGTTCATTCAGAGAGCCCTACCTACCTTAC 3187
Db 61 TGCCTCTTCAAGCAACAGTTCATTCAGAGAGCCCTACCTACCTTAC 120
QY 3188 TGGGTCCTCCAGATGAAGACAGATGCTTTTATTTATTTATTTGGTGTCT 3247
Db 121 TGGGTCCTCCAGATGAAGACAGATGCTTTTATTTATTTATTTGGTGTCT 180
QY 3248 GTGTATTAAAGATCAATATGTAACCACTAGCTCTTTTCACTGACTTAAATAA 3307
Db 181 GTGTATTAAAGATCAATATGTAACCACTAGCTCTTTTCACTGACTTAAATAA 240
QY 3308 CTCACTACTGCTGTGTGATGCTGGGTGTGACTTCTACTGACCGCTAGATAAGCTG 3367
Db 241 CTCACTACTGCTGTGTGATGCTGGGTGTGACTTCTACTGACCGCTAGATAAGCTG 300
QY 3368 TGCCTGCTCCAGAGTGTGGAATTAATTTACATCTGTCCACAGAAAGAAATGTG 3427
Db 301 TGCCTGCTCCAGAGTGTGGAATTAATTTACATCTGTGTCCACAGAAAGAAATGTG 360
QY 3428 TGTGTGAGCAGATGACACATATCTGCTTGTATGAAGAGACTTCTGATTCAGTGTG 3487
Db 361 TGTGTGAGCAGATGACACATATCTGCTTGTATGAAGAGACTTCTGATTCAGTGTG 420
QY 3488 GTTGTGTTATTCATTTGGAATTCATCTTAATCCCATTTGCTATGCTCTAGCA 3547
Db 421 GTTGTGTTATTCATTTGGAATTCATCTTAATCCCATTTGCTATGCTCTAGCA 480
QY 3548 ATTAAGAAATTTCTCAAGTGTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 3604
Db 481 ATTAAGAAATTTCTCAAGTGTGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 535
```

```
RESULT 9
US-09-867-701-3051/c
; Sequence 3051, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867.701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3051
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-3051
Query Match
5.6%; Score 347; DB 10; Length 348;
```


Tue Nov 12 18:03:25 2002

us-09-899-569a-3.mpb

Page 14

Job time : 755 secs

...

QY	1883	CGGTCCTCTTATACCTATTTTCAAAGGAGAAAGCGCTTTCACGCTGACCCCTGACAA	1942
Db	1741	TAGTCTCTATACACCGCTATTTTCAAAGGAAAGGATTTTACGCTGACCCGACAA	1800
QY	1943	AAACCAAGGTCTACCTGAGGAGCCCACTGGGACCGGGCGCTGCATCTCCACCTGTG	2002
Db	1801	AAAACAAGGTCTACCTAAGATTCACCACTGGGACCGTGGCTGCCTCTCTCTAG	1860
QY	2003	TGTCCTGGACATCAGCGTGGCCAGAGACCAAGTGGCCCTGCCTTCTTTAAGAGC	2062
Db	1861	TGTCCTGGACATCAGTGTGGCTTACCAACCAAGTGGCTGTCTGACCGCTTGAAGAAGC	1920
QY	2063	GGAGCGGGGTGTCTGCGCCAGAGGGCGCGCATTCATATATCCAGAGACGCGACCC	2122
Db	1921	GTTCCTGGTGTGGCCCTGCGCCAGTACAGGGCGCTGATATATATATCCAGAGACGCGTCC	1980
QY	2123	GGCGTGAGAGATCTTCAGCCCTGGACGAGATGTGTCTCCCAAGCCAACTTCCACATC	2182
Db	1981	GGCGAGAGAGATCTTCAGCTTGGAGGAGGAAGTCTCTGTAAGCAAGTTTCCACATC	2040
QY	2183	ACAGCTTCTGGGTCAACATCTCTAACTGACAGCCCAAGAGGGAGAGAGCTAGACCTGC	2242
Db	2041	ACAGCTTCTGGGTAACTCTCAATTCACACCCCATAAAGGCAAGCAGTAGATTTCG	2100
QY	2243	TCTTCTGCGTACACTTACCCCAAGGAGCTGGAGATTGACTGTATCTCTCATCGAGGG	2302
Db	2101	TCTTCTGGGTGACCTTACCCCGAGAGCTGTAGATTGGCTG--TCTCATGTGGTGAG	2157
QY	2303	TGGGAGTGGAGTCTTACTGCTGTCTGCCCTCGGGCTCATATTTGCTGTGAAAAA	2362
Db	2158	CAGGAGTGGAGCTCTGCTGCTGTGGCTGGTTCATCATCTGCTTCGAAAAA	2217
QY	2363	AGAAAA 2369	
Db	2218	AAAAAA 2224	
RESULT 2			
LOCUS	AB033417	933 bp	mRNA
DEFINITION	AB033417 Homo sapiens skov Homo sapiens cDNA, mRNA sequence.		EST 18-OCT-1999
ACCESSION	AB033417		
VERSION	AB033417.1	GI:6062545	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	EmkaryOta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 933)		
TITLE	Zhou Y., Kato H. and Wake, N.		
JOURNAL	Biological function unknown		
COMMENT	Unpublished (1999)		
	Contact: Zhou Yong		
	Department of Reproductive Physiology and Endocrinology		
	Medical Institute of Bioregulation, Kyushu University		
	Tsurunihara 4546, Beppu, Oita 874-0838, Japan		
	Tel: 0977-27-1660		
	Fax: 0977-27-1661		
	Email: yzhounetsu@uml.beppu.kyushu-u.ac.jp.		
FEATURES			
source	Location/Qualifiers		
	1..933		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_idb="Homo sapiens skov"		
	/cell_type="ovarian cancer"		
	/cell_line="skov"		
	/note="Buessow, K. et al., 1998, Nucleic Acids Research		
	26(21): 5007-5008"		
BASE COUNT	265 a 251 c 219 g 198 t		
ORIGIN			
Query Match	14.7%; Score 905.8; DB 9; Length 933;		

[illegible]

ACCESSION AN063026
VERSION AN063026.1 GI:6019481
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Zhou, Y., Kato, H. and Wake, N.
TITLE A function unknown partial mRNA, 3' region
JOURNAL Unpublished (1999)
COMMENT Contact: Zhou Yong
Department of Reproductive Physiology and Endocrinology
Medical Institute of Bioregulation, Kyushu University
Tsukumihara 4546, Beppu, Oita 874-0838, Japan
Tel: 0977-27-1660
Fax: 0977-27-1661
Email: yzhou@surumi.beppu.kyushu-u.ac.jp.

FEATURES
source
1..933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="human ovarian cancer cDNA"
/sex="female"
/tissue_type="ovarian cancer"
/cell_line="skov"
/note="Organ: ovary"

BASE COUNT 265 a 251 c 219 g 198 t
ORIGIN

Query Match 14.7%; Score 905.8; DB 10; Length 933;
Best Local Similarly 99.7%; Pred. No. 7.4e-116;
Matches 918; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2215 TCTTACGTCGTGTCGCCCTCGGCTCATCATTTGCTGTGTAAGAAAGAAAGAAAGA 2374
DB 1 TCTTACGTCGTGTCGCCCTCGGCTCATCATTTGCTGTGTAAGAAAGAAAGAAAGA 60

QY 2375 CAACAAGGGGCGCGTGTGGTATCTACATGSCAACATCATATCTAGATGCCGAGGC 2434
DB 61 CAACAAGGGGCGCGTGTGGTATCTACATGSCAACATCATATCTAGATGCCGAGGC 120

QY 2435 AGCCAAAAAGTTTCAAGAAAGGCGAAAGCAATGACTCCCATGTATGCACTATCG 2494
DB 121 AGCCAAAAAGTTTCAAGAAAGGCGAAAGCAATGACTCCCATGTATGCACTATCG 180

QY 2495 AGGACACATGCTATATGGGATGCTGTACAGATTCAGAGGCTCCTCTCTGAGCCAG 2554
DB 181 AGGACACATGCTATATGGGATGCTGTACAGATTCAGAGGCTCCTCTCTGAGCCAG 240.

QY 2555 AGGTGACACCTACCGGCGGTTCCAGGCGCAACATGGGGGTCTGTCTCCCTCCACCCA 2614
DB 241 AGGTGACACCTACCGGCGGTTCCAGGCGCAACATGGGGGTCTGTCTCCCTCCACCCA 300

QY 2615 CCATATGCTCAGGCGCCCACTGCAAAAGTTGGCCACTGAGAGCCACTCTCTGCTCC 2674
DB 301 CCATATGCTCAGGCGCCCACTGCAAAAGTTGGCCACTGAGAGCCACTCTCTGCTCC 360

QY 2675 CTCTGAGTCTGAGAGGAACCGTACACCTTCTCCCATCCCAACAAATGGGATTAAGCA 2734
DB 361 CTCTGAGTCTGAGAGGAACCGTACACCTTCTCCCATCCCAACAAATGGGATTAAGCA 420

QY 2735 GCAAGGACAGACATTCCTTACTGAACACTAGAGCCCATGAGCCAGACAGATTAAC 2794
DB 421 GCAAGGACAGACATTCCTTACTGAACACTAGAGCCCATGAGCCAGACAGATTAAC 480

QY 2795 TTGATCATTCACAGCGCTTGTGCTGATTCATTAAGACGAGGCACTGAGACACCGTCCG 2854
DB 481 TTGATCATTCACAGCGCTTGTGCTGATTCATTAAGACGAGGCACTGAGACACCGTCCG 540

QY 2855 TGTTCCTAACCAAAATCCTTAAGAGGAATTAATACAGAGAACAGCAGAGGTTT 2914
DB 2914 TGTTCCTAACCAAAATCCTTAAGAGGAATTAATACAGAGAACAGCAGAGGTTT 2914

DB 541 TGTTCCTAACCAAAATCCTTAAGAGGAATTAATACAGAGAACAGCAGAGGTTT 600

QY 2915 CCTGACACCGCCCACTTACATTCCTCAGTGACCTCATTTCTAAGGGCAACATTTGAA 2974
DB 601 CCTGACACCGCCCACTTACATTCCTCAGTGACCTCATTTCTAAGGGCAACATTTGAA 660

QY 2975 ATGATGAATTCCAATCTGGATATACATGACAGCTCATGTGCTCTCACTTAAGCTGT 3034
DB 661 ATGATGAATTCCAATCTGGATATACATGACAGCTCATGTGCTCTCACTTAAGCTGT 720

QY 3035 GCGGTAGCCAGCCTGTAATGAGAGAGAGGCGCTGCTACCTAGCATGGGTTGCG 3094
DB 721 GCGGTAGCCAGCCTGTAATGAGAGAGAGGCGCTGCTACCTAGCATGGGTTGCG 780

QY 3095 CAAGCCCTGATTCAGAGTGTAAACAGAGGCTTCCCTCTTACAGACCAAGTTCCAT 3154
DB 781 CGAGCCCTGATTCAGAGTGTAAACAGAGGCTTCCCTCTTACAGACCAAGTTCCAT 840

QY 3155 TCAGAAGAGCCTACCTGAGTCTCTACTCTCAGTGGGTCGCCAGATGAAGACACAT 3214
DB 841 TCAGAAGAGCCTACCTGAGTCTCTACTCTCAGTGGGTCGCCAGATGAAGACACAT 899

QY 3215 GTGCTTTTATTTATTTATTTA 3235
DB 900 GTGCTTTTATTTATTTATTTCA 920

RESULT 4
LOCUS B0962997 928 bp mRNA linear EST 21-AUG-2002
DEFINITION B0962997
ACCESSION B0962997
VERSION B0962997.1 GI:22378463
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgl.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LICM2564 row: g column: 21
High quality sequence stop: 589.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NIH-MGC.42"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOT87; Site: 1: XhoI;
Site: 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library. |"

BASE COUNT 219 a 257 c 238 g 214 t
ORIGIN

Query Match 13.3%: Score 822.4; DB 14; Length 928;
 Best Local Similarity 96.3%: Pred. No. 2.3e-104;
 Matches 886; Conservative 0; Mismatches 26; Indels 8; Gaps 4;

QY 308 CTATGCGACTGCTAGGGGTTCTGCTGGTGCGGGCGGCTGCGCGCGGGGCGAAG 367
 |||||
 DB 1 CTATGCGACTGCTAGGGGTTCTGCTGGTGCGGGCGGCGGCTGCGCGCGGGGCGAAG 60
 QY 368 CTTTGAAGTTCTCTGCGCAGAGAAAGCAATTACATTCTCATTAAGCTGGGACCC 427
 |||||
 DB 61 CTTTGAAGTTCTCTGCGCAGAGAAAGCAATTACATTCTCATTAAGCTGGGACCC 120
 QY 428 CGACTGTGCGCAAAACCTGTATCATGCTATTCTTAAAGACATTAACATGTTGT 487
 |||||
 DB 121 CGACTGTGCGCAAAACCTGTATCATGCTATTCTTAAAGACATTAACATGTTGT 180
 QY 488 CCATCAAGTCTGAGAAAGAAATAGTCTTACCTTTAGTCCAGAGTCCCTGAGAACTAGT 547
 |||||
 DB 181 CCATCAAGTCTGAGAAAGAAATAGTCTTACCTTTAGTCCAGAGTCCCTGAGAACTAGT 240
 QY 548 TTGTCTATAGATCCAGAAATATTGACTGTATGTACAGGCCCATGCTCTTTGGGAGG 607
 |||||
 DB 241 TTGTCTATAGATCCAGAAATATTGACTGTATGTACAGGCCCATGCTCTTTGGGAGG 300
 QY 608 TTACAGTTCAGCCCTGACATGTTGTTGCTTACCTCAACCTCAAGAACTTTCATCTGGGATG 667
 |||||
 DB 301 TTACAGTTCAGCCCTGACATGTTGTTGCTTACCTCAACCTCAAGAACTTTCATCTGGGATG 360
 QY 668 TCMAAGCTCATTAAGAGCATCGGTTTAGAGCTCAGTTCATCCCTCCCTGAGGACAGA 727
 |||||
 DB 361 TCMAAGCTCATTAAGAGCATCGGTTTAGAGCTCAGTTCATCCCTCCCTGAGGACAGA 420
 QY 728 TCGGTCCGGGTAGAGACTGCCAGAGGATCATCTCATTCACAGCGGCCGAATGATG 787
 |||||
 DB 421 TCGGTCCGGGTAGAGACTGCCAGAGGATCATCTCATTCACAGCGGCCGAATGATG 480
 QY 788 CGACGCTGTCAGAGTCCGAACCTTCTGAGCAATGGCACTGTGTCCCGGATCAATATGC 847
 |||||
 DB 481 CGACGCTGTCAGAGTCCGAACCTTCTGAGCAATGGCACTGTGTCCCGGATCAATATGC 540
 QY 848 AAGAAGAGTGAATAATGCGCTTACACCTCCATGTTCCACCCAGAAATGTCTCGGCT 907
 |||||
 DB 541 AAGAAGAGTGAATAATGCGCTTACACCTCCATGTTCCACCCAGAAATGTCTCGGCT 600
 QY 908 TCAGCATTCGCAAAACCGCTCATCTATTAAGAGTGTGATCATCGAGTGTGTGAGG 967
 |||||
 DB 601 TCAGCATTCGCAAAACCGCTCATCTATTAAGAGTGTGATCATCGAGTGTGTGAGG 660
 QY 968 GTGAAGGCTCAGAACCCGATGTGCGCACTACCCAGAAAGGCTCCCTGAGGATGAGC 1027
 |||||
 DB 661 GTGAAGGCTCAGAACCCGATGTGCGCACTACCCAGAAAGGCTCCCTGAGGATGAGC 720
 QY 1028 TCATGACGTGGAGATTGTTGCTGCTGCAACCTGCGGGCCAGCGTCTCTTCTCAACT 1087
 |||||
 DB 721 TCATGACGTGGAGATTGTTGCTGCTGCAACCTGCGGGCCAGCGTCTCTTCTCAACT 780
 QY 1088 TCACCTCTCCAACTGTGAGAGAGAGA-GGAGCGGGTGAATATACATCTCCGGGC-TC 1145
 |||||
 DB 781 TCACCTCTCTCCAACTGTGAGAGAGAGAGAGCGGGTGAATATACATCTCCGGGC-TC 840
 QY 1146 GAGCAGCAACCCCGA--GGTGTTCAGCT----GGAGGCAAGCAGCCGGAACATGGC 1199
 |||||
 DB 841 GAGCAGCAACCCCGAAGGTTGTTCAAGCTGGAGGAGAAAGCAGCGCTGTGGACCATGGC 900
 QY 1200 GGGCAACTTCAACTCTCTC 1219
 |||||
 DB 901 GGGCAACTTCAACTCTCTC 920

RESULT 5
 B0687577
 LOCUS B0687577 868 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT_8343949 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6248364
 5', mRNA sequence.
 ACCESSION B0687577
 VERSION B0687577.1 GI:21812893
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 868)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM2388 row: b column: 13
 High quality sequence stop: 702.
 Location/Qualifiers
 1..868
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6248364"
 /clone_lib="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 214 a 238 c 206 g 210 t
 ORIGIN

Query Match 13.2%: Score 813.6; DB 14; Length 886;
 Best Local Similarity 98.1%: Pred. No. 4e-103;
 Matches 844; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 5065 CAAGACCTCTGAAACCTTCCAGAGATAGATATTAAATGATGCTCCCTGGCTGGCC 5124
 |||||
 DB 1 CAAGACCTCTGAAACCTTCCAGAGATAGATATTAAATGATGCTCCCTGGCTGGCC 60
 QY 5125 TATGGACCTTTCCTTGTGAAGTCTGTTCTGCGCCAGTGACCTTGCTGTGAGC 5184
 |||||
 DB 61 TATGGACCTTTCCTTGTGAAGTCTGTTCTGCGCCAGTGACCTTGCTGTGAGC 120
 QY 5185 CGAGATGCTGACCTTCATTAAGGCGCAAGAGGCGCTGGCGTTCCTTCCCTACTGAA 5244
 |||||
 DB 121 CGAGATGCTGACCTTCATTAAGGCGCAAGAGGCGCTGGCGTTCCTTCCCTACTGAA 180
 QY 5245 GAGCCCTTATTTGAATTCACTGTGTGAGAGCCCTGAGCCCTTCATTTTCGACATTCGCCAAC 5304
 |||||
 DB 181 GAGCCCTTATTTGAATTCACTGTGTGAGAGCCCTGAGCCCTTCATTTTCGACATTCGCCAAC 240
 QY 5305 CTCGCCAGCCCTTCCAAAGAGAGAGTGTGAGCCCTGAGATTCACCCCAAGTGGGATTGGCC 5364
 |||||
 DB 241 CTCGCCAGCCCTTCCAAAGAGAGAGTGTGAGCCCTGAGATTCACCCCAAGTGGGATTGGCC 300
 QY 5365 TTCTTAAGCTGTGCTACTGTACCATTCACCGACATCATCTGTGCTGCAAGGACACAC 5424
 |||||
 DB 301 TTCTTAAGCTGTGCTACTGTACCATTCACCGACATCATCTGTGCTGCAAGGACACAC 360
 QY 5425 GTGGGCAATTTCTCTCACTGAGGAGCTCAAAATCTCTGGCAAGTGTGAGGCTCCGAGAA 5484
 |||||

Db 361 GTGGCATTTCCTTCAACGTGAGGGCTCAAAACTCTTGACAACTCTGCTCCTAGAGA 420
 QY 5485 CCACTATTCTCTGAGCTGTGCTCACTGAGAGGGCCAGCTGAGAACCTGCTCTT 5544
 Db 421 CCACTATTCTCTGAGCTGTGCTCACTGAGAGGGCCAGCTGAGAACCTGCTCTT 480
 QY 5545 TTCTTTAAAGCCCGCCCACTTACATTAATAACATTTCAGGGCTCAGTGAACAGTGAAG 5604
 Db 481 TTCTTTAAAGCCCGCCCACTTACATTAATAACATTTCAGGGCTCAGTGAACAGTGAAG 540
 QY 5605 TGCATTGTTGAAGCTTACATGATGAGCCAGCCCTGCTCAATCCAGCTGGCTGCTGAGC 5664
 Db 541 TGCATTGTTGAAGCTTACATGATGAGCCAGCCCTGCTCAATCCAGCTGGCTGCTGAGC 600
 QY 5665 CTACAGAGAAAGCCAGCGCATGACAGAGCTGCTCTAATGCTGTGTCATTGACAGAG 5724
 Db 601 CTACAGAGAAAGCCAGCGCATGACAGAGCTGCTCTAATGCTGTGTCATTGACAGAG 660
 QY 5725 GGAAGGCTTCAGAGAGATCACTGAGGACAGACAGACAGCCAGGACATGGCTTGG 5784
 Db 661 GGAAGGCTTCAGAGAGATCACTGAGGACAGACAGACAGCCAGGACATGGCTTGG 720
 QY 5785 TAAAGGTTACAGAGCTGTGCTGATGCTGACAGCTTCACTGAGAAATATTATTCA 5844
 Db 721 TAAAGGTTACAGAGCTGTGCTGATGCTGACAGCTTCACTGAGAAATATTATTCA 780
 QY 5845 TTGCAGATCTTTTAAAGTGG-CATTTTATTCAATTCCTGTCG-TTTAAATAAACAATG 5902
 Db 781 TTGCAGATCTTTTAAAGTGGCCATTTTATTCAATTCCTGTCGCTTTTAAAAAACAAG 840
 QY 5903 TACCAAAAACAAGTATCA 5922
 Db 841 TACCCCAAAAAA 860

RESULT 6
 LOCUS B0691792 1008 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8034267 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208223
 ACCESSION B0691792
 VERSION B0691792.1 GI:21817108
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 1008)
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.linl.gov
 Plate: LHCW365 row: 1 column: 24
 High quality sequence start: 27
 High quality sequence stop: 744.
 Location/Qualifiers
 1..1008

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6208223"
 /clone_lib="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOPB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming."

BASE COUNT 238 a 238 c 224 g 246 t 62 others
 ORIGIN
 Query Match 13.1%; Score 810.4; DB 14; Length 1008;
 Best Local Similarity 98.6%; Pred. No. 1e-102;
 Matches 847; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
 Note: this is a NIH-MGC library.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 QY 5289 CTGACATTCCTCCCACTCCAGCCCTTCCAAAGAGAGTGGCTGATTCAC 5348
 Db 69 CTCGACATTCCTCCCACTCCAGCCCTTCCAAAGAGAGTGGCTGATTCAC 128
 QY 5349 CAAGGTGGATTTGGCTCTTCTTAAAGCTGCTACTGTGCACCATCAGCATCTCTG 5408
 Db 129 CAAGGTGGATTTGGCTCTTCTTAAAGCTGCTACTGTGCACCATCAGCATCTCTG 188
 QY 5409 CTTGCAAGAGACACACAGTGGCCATTTCTTCAACTAGAGGCTCAAACTCTGACAG 5468
 Db 189 CTTGCAAGAGACACACAGTGGCCATTTCTTCAACTAGAGGCTCAAACTCTGACAG 248
 QY 5469 TTGCTGCTCCTGAGACAGTATTCTTGAAGCTGTGCTCACTGAGAGGGCCAGCTG 5528
 Db 249 TTGCTGCTCCTGAGACAGTATTCTTGAAGCTGTGCTCACTGAGAGGGCCAGCTG 308
 QY 5529 AGGACCTGCTGCTCTTCTTAAAGCCAGCCCACTTCAATTAATTCAGGCTG 5588
 Db 309 AGGACCTGCTGCTCTTCTTAAAGCCAGCCCACTTCAATTAATTCAGGCTG 368
 QY 5589 ACTGGAACAGTGAAGTCCATTTGTTGAAGCTTCTGATGAGCCAGCTGCTATCC 5648
 Db 369 ACTGGAACAGTGAAGTCCATTTGTTGAAGCTTCTGATGAGCCAGCTGCTATCC 428
 QY 5649 ACGTGTCTGCTGCTTCTGATGAGAGAGCCAGCTGCTGCTGCTGCTGCTG 5708
 Db 429 ACGTGTCTGCTGCTTCTGATGAGAGAGCCAGCTGCTGCTGCTGCTGCTGCTG 488
 QY 5709 GGTGATTCAGACAGAGGAGGAGGCTCTCAAGAGAGTCACTGGGACAGACAGCCCA 5768
 Db 489 GGTGATTCAGACAGAGGAGGAGGCTCTCAAGAGAGTCACTGGGACAGACAGCCCA 548
 QY 5769 CCGGACATGAGCTGTGTTAAAGTTAGCAGACTGTGTGTGATCTGCACTGCTGCT 5828
 Db 549 CCGGACATGAGCTGTGTTAAAGTTAGCAGACTGTGTGTGATCTGCACTGCTGCT 608
 QY 5829 GGAATATATTATTCATTCAGATACATTTTATGAGTGCATTTTATTCATTTCTGCT 5888
 Db 609 GGAATATATTATTCATTCAGATACATTTTATGAGTGCATTTTATTCATTTCTGCT 668
 QY 5889 TAAATTAACAAGTACCAAAAAACAAGTATCAAGCTTAACTGGTGGCTGCTGCT 5948
 Db 669 TAAATTAACAAGTACCAAAAAACAAGTATCAAGCTTAACTGGTGGCTGCTGCT 728
 QY 5949 CCCTGTGTTAGTGAAGAGCCCGGTTTCCAGTGTGATGAGTGAAGCTGACAGCATGAG 6008
 Db 729 CCCTGTGTTAGTGAAGAGCCCGGTTTCCAGTGTGATGAGTGAAGCTGACAGCATGAG 788
 QY 6009 CTCAGCAGATGCTGCTTAAATTTGTGATATACAGAAAGCCAGGCTTGGGATACAGN 6067
 Db 789 CTCAGCAGATGCTGCTTAAATTTGTGATATACAGAAAGCCAGGCTTGGGATACAGN 848
 QY 6068 TTCTTTCTCTTCAATTTGATGAGCCG-TGCACATGTGTGAAGAGATGTTTGTGCGG-AAA 6125
 Db 849 NTCTTTCTCTTCAATTTGATGAGCCGNNCAGTGTGTGAAGAGATGTTTGTGCGGAAA 908
 QY 6126 TAAATAATATAGTCTTGA 6144
 Db 909 TAAATAATATAGTCTTGA 927


```
RESULT 7
B0951236 984 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT.8817950 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6422238
DEFINITION 5', mRNA sequence.
ACCESSION B0951236
VERSION B0951236.1 GI:22366714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 984)
NIH-MGC http://mhc.nci.nih.gov/.
REFERENCE National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS Unpublished (1999)
TITLE Contact: Robert Strausberg, Ph.D.
JOURNAL Email: cgapbs-remail.nih.gov
COMMENT Tissue Procurement: DCTD/DP/GenZdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM2598 row: 0 column: 07
High quality sequence start: 6
High quality sequence stop: 719.
Location/Qualifiers
1..984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6422238"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."
BASE COUNT 238 a 301 c 247 g 197 t 1 others
ORIGIN
Query Match 12.9% Score 794.4; DB 14; Length 984;
Best Local Similarity 96.7%; Pred. No. 1.6e-100;
Matches 832; Conservative 0; Mismatches 26; Indels 2; Gaps 2;
QY 951 CGAGTGTGTGTGAGGGTGAAGGCTCAGCAACCTGATGTCGCCAATACCCAGAGG 1010
DB 18 CGAGGCTGTGTGAGGGTGAAGGCTCAGCAACCTGATGTCGCCAATACCCAGAGG 77
QY 1011 CTTCCCTGAGAGTACTATGACGTGGGCGAGTTTGCTTCCTGCGACACTGGGGCCAG 1070
DB 78 CTTCCCTGAGAGTACTATGACGTGGGCGAGTTTGCTTCCTGCGACACTGGGGCCAG 137
QY 1071 CGTCTCTTCTCACTTCAACTCTTCACACTGTGAGAGAGAGAGAGGGGTTGAATA 1130
DB 138 CGTCTCTTCTCACTTCAACTCTTCACACTGTGAGAGAGAGAGAGGGGTTGAATA 197
QY 1131 CTACATCCGGGGCTCACCACCAACCCGAGGTTCACAGCTGGAGACAAGCCCTGG 1190
DB 198 CTACATCCGGGGCTCACCACCAACCCGAGGTTCACAGCTGGAGACAAGCCCTGG 257
QY 1191 GAACATGGGGGGGAATCTTCAACTCTCTCTGCAAGGCTGAGACCAAGATCCCAAGTCC 1250
DB 258 GAACATGGGGGGGAATCTTCAACTCTCTCTGCAAGGCTGAGACCAAGATCCCAAGTCC 317
QY 1251 AGGATCTCTCCGGCTCAAGTTTGTGTCACAACATCCACAAAATGAAGCAATAA 1310
```

```
DB 318 AGGATCTCTCCGGCTCAAGTTTGTGTCACAACATCCACAAAATGAAGCAATAA 377
QY 1311 AATCTACGTGTGACTTGAATGATGAGCAGCAGCCATCTACTACCATGACCGAGGCC 1370
DB 378 AATCTACGTGTGACTTGAATGATGAGCAGCAGCCATCTACTACCATGACCGAGGCC 437
QY 1371 CGTCAACAGAGCGGCAAGTTTGCTCCCTGGCTTTCTGCTGCTGAATCTGGAGCTG 1430
DB 438 CGTCAACAGAGCGGCAAGTTTGCTCCCTGGCTTTCTGCTGCTGAATCTGGAGCTG 497
QY 1431 CAGTAGCAACCTCAACCTTGACATCTGGCTCCAAACACAAATCTCTTCTTGATGA 1490
DB 498 CAGTAGCAACCTCAACCTTGACATCTGGCTCCAAACACAAATCTCTTCTTGATGA 557
QY 1491 TCTGACACGTCTGTGATGTAATGTGAAAAACCATTAAGTGCACAGACCGGTACTG 1550
DB 558 TCTGACACGTCTGTGATGTAATGTGAAAAACCATTAAGTGCACAGACCGGTACTG 617
QY 1551 CCAAGGAATCTCTACTCTCCAGGTCGCGCAGGACATCTCCACCTGCTGGAGCT 1610
DB 618 CCAAGGAATCTCTACTCTCCAGGTCGCGCAGGACATCTCCACCTGCTGGAGCT 677
QY 1611 GCATGACTTCTCTGGAAGCTGTGTGCCCCAAGAGACAGGCTGAGTGTGCC 1670
DB 678 GCATGACTTCTCTGGAAGCTGTGTGCCCCAAGAGACAGGCTGAGTGTGCC 737
QY 1671 AGCCCAAGACCTGACGACGATACACAGAGAAAGCCCTGCAACACGAC-TTACGTAAC 1729
DB 738 AGCCCAAGACCTGACGACGATACACAGAGAAAGCCCTGCAACACGATTAACCTAC 797
QY 1730 TCGTGGCCATGCGCATTCGCCAGGC-AGACCTGACTTGCGCTCTGCCCCGGGAGGC 1788
DB 798 TCGGGGCGACTGATCCACGACGAGGACCTGACTTCGGCTCTTTGTGCCGGGGA 857
QY 1789 TCTATCAAGCAGATCCAGT 1808
DB 858 GAGCTCTATCAACCAAGAT 877
RESULT 8
B0889198 871 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT.8072366 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6084315
DEFINITION 5', mRNA sequence.
ACCESSION B0889198
VERSION B0889198.1 GI:22281212
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 871)
NIH-MGC http://mhc.nci.nih.gov/.
REFERENCE National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS Unpublished (1999)
TITLE Contact: Robert Strausberg, Ph.D.
JOURNAL Email: cgapbs-remail.nih.gov
COMMENT Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM2315 row: 0 column: 04
High quality sequence stop: 707.
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6084315"
/clone_lib="NIH_MGC_110"
```

```

/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/node="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library."
BASE COUNT      211 a      247 c      176 g      237 t
ORIGIN
Query Match      12.7%: Score 784.8; DB 14; Length 871;
Best Local Similarity 99.5%; Pred. No. 3.7e-99;
Matches 808; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 3983 GATATATCTTCATCTAAAGATCCCTCATCTCGGAGAGAGCCCTTGGCATGCAGAC 4042
DB 1 GATATATCTTCATCTAAAGATCCCTCATCTCGGAGAGAGCCCTTGGCATGCAGAC 60
QY 4043 AACATAGCCACAGGTGGGATTAGACAGACATCTTGGGTCGTATTCTGCCTA 4102
DB 61 AACATAGCCACAGGTGGGATTAGACAGACATCTTGGGTCGTATTCTGCCTA 120
QY 4103 CCACACCTCTGCACTGACTCCACAGAGAGGCTACAAATGATCTGGCGACAGGG 4162
DB 121 CCACACCTCTGCACTGACTCCACAGAGAGGCTACAAATGATCTGGCGACAGGG 180
QY 4163 ATGTTTGTAGCTGGGAGCTCTACACATT-AAAAAACCCAGATCAAAATCTGG 4221
DB 181 ATGTTTGTAGCTGGGAGCTCTACACATTAAAAAACCCAGATCAAAATCTGG 240
QY 4222 CCATGCTGGGCTCACATTCTCACCTAGCAACAATGGCTGGAGTGGGACAGCTGTG 4281
DB 241 CCATGCTGGGCTCACATTCTCACCTAGCAACAATGGCTGGAGTGGGACAGCTGTG 300
QY 4282 CCTTTAGAGGGGTGCTCATTACACAGGTACACAGCCACACTACGCTTACTCTT 4341
DB 301 CCTTTAGAGGGGTGCTCATTACACAGGTACACAGCCACACTACGCTTACTCTT 360
QY 4342 CCCCAATGAGGCTAAGGTTTGTCTACTGATCAATGCCCGCAGAGTGGCATTAAT 4401
DB 361 CCCCAATGAGGCTAAGGTTTGTCTACTGATCAATGCCCGCAGAGTGGCATTAAT 420
QY 4402 GTATATGAAAAAGAAAGACTGGGATTAAATCTTAATCAGGTAGTGAACATGACCAAT 4461
DB 421 GTATATGAAAAAGAAAGACTGGGATTAAATCTTAATCAGGTAGTGAACATGACCAAT 480
QY 4462 GTGTGCTCACATTACCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 4521
DB 481 GTGTGCTCACATTACCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 540
QY 4522 GAGACAGGATCTCATCTGTTGCTAGGCTAGGAGTGGAGTGGGCAATCTGGGCTACTG 4581
DB 541 GAGACAGGATCTCATCTGTTGCTAGGCTAGGAGTGGAGTGGGCAATCTGGGCTACTG 600
QY 4582 CAACCTGCTGCTCTGGGCTCAAGCAATTCCTCCACTCAGCTCCCAAAATAGTGGGAT 4641
DB 601 CAACCTGCTGCTCTGGGCTCAAGCAATTCCTCCACTCAGCTCCCAAAATAGTGGGAT 660
QY 4642 CACTGGCAACAACACATGCGCAGCTAATTTTGTATTTTGTAGAGACAGGGTTTCAC 4701
DB 661 CACTGGCAACAACACATGCGCAGCTAATTTTGTATTTTGTAGAGACAGGGTTTCAC 720
QY 4702 CATTTTGGCCAGGCTGTGCTCAACCTCTGGGCTCAAGCAATCTCTCTGCTGGCTCC 4761
DB 721 CATTTTGGCCAGGCTGTGCTCAACCTCTGGGCTCAAGCAATCTCTCTGCTGGCTCC 780
QY 4762 C-AAAGTGTGGATATACAGATGTGAGCCACC 4792
DB 781 CAAAAGTGTGGATATACAGATGTGAGCCACC 812

```

```

RESULT 9
BE877551
LOCUS
DEFINITION
601486503F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3888768 5',
mRNA sequence.
ACCESSION
BE877551
VERSION
BE877551.1 GI:10326327
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1150)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: ggaups-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHM9669 row: f column: 01
High quality sequence stop: 724.
FEATURES
location/qualifiers
source
1..1150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3888768"
/clone_1db="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/node="Organ: lung; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      259 a      329 c      312 g      250 t
ORIGIN
Query Match      12.1%: Score 742.8; DB 12; Length 1150;
Best Local Similarity 96.5%; Pred. No. 2e-93;
Matches 770; Conservative 0; Mismatches 27; Indels 1; Gaps 1;
QY 2412 CATCAATACAGATGCCGAGGACCAAAATTTTCAGAAAGGCGAAAGCAATGA 2471
DB 1 CATCAATACAGATGCCGAGGACCAAAATTTTCAGAAAGGCGAAAGCAATGA 60
QY 2472 CTCCCATGTGATGACAGTATCGAGACACCATGATATGGGATCGTACAGGATTC 2531
DB 61 CTCCCATGTGATGACAGTATCGAGACACCATGATATGGGATCGTACAGGATTC 120
QY 2532 CAGCGGCTCTCTCTGAGCCAGAGGTGACACTTACCGGCGCTTCCAGGCGACCATGG 2591
DB 121 CAGCGGCTCTCTCTGAGCCAGAGGTGACACTTACCGGCGCTTCCAGGCGACCATGG 180
QY 2592 GGTGTGTCCTCCCTCCGACACCATATGCTCCAGGGCCCCCACTGAAAGTTGGCCAC 2651
DB 181 GGTGTGTCCTCCCTCCGACACCATATGCTCCAGGGCCCCCACTGAAAGTTGGCCAC 240
QY 2652 TGAGGAGCCACCTCTGCTGCTCTCTCTGAGTGTGAGATGAACCGTACACTTCCCA 2711
DB 241 TGAGGAGCCACCTCTGCTGCTCTCTCTGAGTGTGAGATGAACCGTACACTTCCCA 300
QY 2712 TCCCAACAATGGGATTAAGCAGACAGACATTCCTTACTGAACACTCAGGA 2771
DB 301 TCCCAACAATGGGATTAAGCAGACAGACATTCCTTACTGAACACTCAGGA 360
QY 2772 GCCCATGAGGACAGCAATTAATCTTGAATCCATTTCAGAGCGTTTGTGAGTTTATAAG 2831

```

Db 361 GCCCATGAGCCAGCAAAATACCTGATCCATTCAGAGCGCTTGTGATTCATAAG 420
QY 2832 CAGGGACTGAGACACCCCTCGTGTCTTACACAGAAATCTTAAGAGAGAAATTA 2891
Db 421 CAGGGGACAGACACCCCTCGTGTCTTACACAGAAATCTTAAGAGAGAAATTA 480
QY 2892 CAGGAAGAACAGAGAGGTTTTCCTGAGACACCGCAACTTCATTCAGTGCAGTGCAC 2951
Db 481 CAGAGAGAACAGAGAGGTTTTCCTGAGACACCGCAACTTCATTCAGTGCAGTGCAC 540
QY 2952 ATTCTAAGGCAAGACATTGAAATGATGATATTCATCTGATACAGTATGACAGCTC 3011
Db 541 ATTCTAAGGCAAGACATTGAAATGATGATATTCATCTGATACAGTATGACAGCTC 600
QY 3012 ATGTGCTCTCAACTTAAAGCTGTGCGGTTAAGCAGCTGTATGAGAGAGAGAGCTTC 3071
Db 601 ATGTGCTCTCAACTTAAAGCTGTGCGGTTAAGCAGCTGTATGAGAGAGAGAGCTTC 660
QY 3072 AGTCACCTAGCAATAGGTTGACAGCAAGCCCTGATTCAGAGTGTAAACAGAGAGCTTGC 3131
Db 661 AGTCACCTAGCAATAGGTTGACAGCAAGCCCTGATTCAGAGTGTAAACAGAGAGCTTGC 720
QY 3132 CTCTTACGAGCAACAGATTCATTCAGAGAGAGCTGACCTGA-GGTCCCTACTCTCACTGG 3190
Db 721 TCTTCAGGGACACCACTTCAATTTCCAGGAGAGCTTAAGTGAAGGATCTCACTGG 780
QY 3191 GGTCCCGAGAGTGAAC 3208
Db 781 GGTCCCGAGAGTGAAC 798

RESULT 10
Bg119563 851 bp mRNA linear EST 14-MAR-2001
LOCUS 602451816P1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:459054 5',
DEFINITION mRNA sequence.
ACCESSION Bg119563
VERSION Bg119563.1 GI:13326109
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 851)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: gsa@ncl.nih.gov
Tissue Procurement: DCTP/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Place: LNCMI325 row: 0 column: 03
High quality sequence stop: 699.
Location/Qualifiers
1. 851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:459054"
/clone_1lb="NIH_MGC_14"
/issue_type="renal cell adenocarcinoma"
/note="Organ: kidney; Vector: pCMV; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit

BASE COUNT 181 a 236 c 240 g 194 t
ORIGIN
Query Match 11.5%; Score 709.4; DB 12; Length 851;
Best Local Similarity 94.4%; Pred. No. 9.4e-89;
Matches 779; Conservative 0; Mismatches 41; Indels 5; Gaps 4;
QY 148 GGGCGGGGCTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 207
Db 2 GGGCGGGGCTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 60
QY 208 GGGCGGGGCTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 267
Db 61 GGGCGGGGCTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 120
QY 268 AGGCGTCCCGGAGTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 327
Db 121 AGGCGTCCCGGAGTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 328 CTGCTGTGGGTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 387
Db 181 CTGCTGTGGGTGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 240
QY 388 CGAGAAAGCAACATTAAGTTCATTAAGTTCATTAAGTTCATTAAGTTCATTAAGT 447
Db 241 CGAGAAAGCAACATTAAGTTCATTAAGTTCATTAAGTTCATTAAGTTCATTAAGT 300
QY 448 TGTTCATCGTCAATTTCTTAAGACATTAAGACATTAAGACATTAAGACATTAAG 507
Db 301 TGTTCATCGTCAATTTCTTAAGACATTAAGACATTAAGACATTAAGACATTAAG 360
QY 508 ATAGCTTTACCTTTAGCTCCAGAGTCTGAGAGATCACTTTGTATGATGATCAGAAA 567
Db 361 ATAGCTTTACCTTTAGCTCCAGAGTCTGAGAGATCACTTTGTATGATGATCAGAAA 420
QY 568 AATATTGACGTATATGACGGCCCATGCTTTGGGGAGGTTGAGCTTCAAGCTCGACA 627
Db 421 AATATTGACGTATATGACGGCCCATGCTTTGGGGAGGTTGAGCTTCAAGCTCGACA 480
QY 628 TCGTTTGGCCCTACCTCAACAGAACTTTCATGAGTCAAGCTCAAGTCAAGAGATC 687
Db 481 TCGTTTGGCCCTACCTCAACAGAACTTTCATGAGTCAAGCTCAAGTCAAGAGATC 540
QY 688 GGTTCAGAGTCAAGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 747
Db 541 GGTTCAGAGTCAAGTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 600
QY 748 CCAGAGGAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 807
Db 601 CCAGAGGAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 660
QY 808 ACCTTTCGACGAA-TGGCACTGTGCC-GGATCAAGATGCAAGAG--AGTGAAT 863
Db 661 ACCTTTCGACGAAATTTGGCACTGTGCCGGGATCAAGATGCAAGAGAGTGAAT 720
QY 864 GGCCTTACCTCCCATGTTTCCAGCCAGAAATGTCCTCGGCTTCAGATTGCAACCG 923
Db 721 GGCCTTACCTCCCATGTTTCCAGCCAGAAATGTTTCCGGTTCAGATTGCAACCG 780
QY 924 CTATCATATAAAGCTGTGATCATCATCATCATCATCATCATCATCATCATCATCAT 968
Db 781 GTATTTTATAAAGTTGTCTATCGCTTGGTGAAGG 825

RESULT 11
Bg119556 779 bp mRNA linear EST 30-JAN-2001
LOCUS 602347259P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4442127 5',
DEFINITION mRNA sequence.
ACCESSION Bg119556
VERSION Bg119556.1 GI:12613062
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
AUTHORS 1 (bases 1 to 779)
TITLE NIH-MGC http://mgc.ncl.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LHAM0214 row: f column: 16
High quality sequence stop: 700.

FEATURES
Location/Qualifiers
1..779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4442127"
/clone_1lb="NIH-MGC_90"
/issue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 185 a 215 c 161 g 218 t

ORIGIN

Query Match 11.2%; Score 691.2; DB 12; Length 779;
Best Local Similarity 96.1%; Pred. No. 3.2e-86;
Matches 742; Conservative 0; Mismatches 23; Indels 7; Gaps 3;

QY 3977 ATCCAGATATATCTCCATCTAAAGATCCCTTCATCTCTGGAAGACCTTTGGCATG 4036
|||||
Db 1 ATCCAGATATATCTCCATCTAAAGATCCCTTCATCTCTGGAAGACCTTTGGCATG 60

QY 4037 CAAGACACATAGCCACAGGTGGGATTTAGACACGACATCTTTGGGGTCTGTATTC 4096
|||||
Db 61 CAAGACACATAGCCACAGGTGGGATTTAGACACGACATCTTTGGGGTCTGTATTC 120

QY 4097 TGCTTACACACCTTCCTGCACAGTACCTCCACAGAGAGGCTCAAAATGATCTGGCGC 4156
|||||
Db 121 TGCTTACACACCTTCCTGCACAGTACCTCCACAGAGAGGCTCAAAATGATCTGGCGC 180

QY 4157 ACAGGATGTTTTGTTAGCTTGGGACTCTAACACTT-AAAAAACCCAGATCAGAG 4215
|||||
Db 181 ACAGGATGTTTTGTTAGCTTGGGACTCTAACACTTAAAAAACCCAGATCAGAG 240

QY 4216 ATCTGCCATGCTGGGCTCACAATCTCAGCTAGCAACAACCTGGCTGGAGTGGACCA 4275
|||||
Db 241 ATCTGCCATGCTGGGCTCACAATCTCAGCTAGCAACAACCTGGCTGGAGTGGACCA 300

QY 4276 GCTGTGCTTTAGAAAGGGGTGCACCTTCACAGTACACACACCCACACTAGCCCTA 4335
|||||
Db 301 GCTGTGCTTTAGAAAGGGGTGCACCTTCACAGTACACACACCCACACTAGCCCTA 360

QY 4336 TCACTTCCCAATAGAGCTTAAGTGTGTTTCTACTGATCAATGCCCTGCAGGTGCA 4395
|||||
Db 361 TCACTTCCCAATAGAGCTTAAGTGTGTTTCTACTGATCAATGCCCTGCAGGTGCA 420

QY 4396 TTTATTTAATGAAAAAAGAACTGGGATTAATCTCTAATCAGTGAATGACCATGAG 4455
|||||
Db 421 TTTATTTAATGAAAAAAGAACTGGGATTAATCTCTAATCAGTGAATGACCATGAG 480

QY 4456 ACCAATGTGCTCACAATTACCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4515
|||||

Db 481 ACCAATGTGCTCACAATTACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTC--TTTTTTT 538

QY 4516 TAATGTGAGACAGATCTCATTTCTGTTGCTTAGCTGGAGTGCAGTGGCGAATCTCGGC 4575
|||||

Db 539 TAATGTGAGACAGATCTCATTTCTGTTGCTTAGCTGGAGTGCAGTGGCGAATCTCGGC 598

QY 4576 TCACTGCAACCTTGCCCTCTGGGCTCAGCAATTTCCACCTCAGCTCCCAATAGC 4635
|||||

Db 599 TCACTGCAACCTTGCCCTCTGGGCTCAGCAATTTCCACCTCAGCTCCCAATAGC 658

QY 4636 TGGATGCTGCGCAAAACCAACCACTAATTTGATTTGTAGAGCAGG 4695
|||||

Db 659 TGGATGCTGCGCAAAACCAACCACTGAGCAG--TAATTTGATTTGTACAGACCGG 714

QY 4696 TTTACATGTTGGCCAGGCTGTCTCAACCTCTGAGGCTCAGCAATCTC 4747
|||||

Db 715 GTTCCCATGTTTGGCCAGGTTGTCTCAACCTCTGAGGCTCAGCAATCTTC 766

RESULT 12
BQ950224 977 bp mRNA linear EST 21-AUG-2002
LOCUS BQ950224
DEFINITION AGENCOURT_8801117 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311019
5', mRNA sequence.
ACCESSION BQ950224
VERSION BQ950224.1 GI:22365702
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 977)
TITLE NIH-MGC http://mgc.ncl.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Resgen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: LHAM13734 row: e column: 04
High quality sequence start: 7
High quality sequence stop: 496.

FEATURES
Location/Qualifiers
1..977
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6311019"
/clone_1lb="NIH-MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH-MGC Library."

BASE COUNT 242 a 269 c 190 g 274 t 2 others

ORIGIN

Query Match 11.2%; Score 688.8; DB 14; Length 977;
Best Local Similarity 95.1%; Pred. No. 6e-86;
Matches 742; Conservative 0; Mismatches 34; Indels 4; Gaps 3;

QY 4214 AGATCGGCATGCTGGGGCTCACATTTCTACGACAAACAATGGCTGGAGCTGGCAG 4273
|||||

Db 8 ATNNCTGGCATGCTGGGGCTCACATTTCTACGACAAACAATGGCTGGAGCTGGCAG 67

QY 4274 CAGCTGCGCTTTAGAAAGGGGTGCACCTTCACAGGTCACACAGCCCACTACAGGCC 4333
|||||

Db 68 CAGCTGCGCTTTAGAAAGGGGTGCACCTTCACAGGTCACACAGCCCACTACAGGCC 127

```

OY 4334 TATCACTCCACATGAGGCTAGTGTGTTCTACTGATCAATGCCCTGACGTTG 4393
    |||||||
DB 128 TATCACTCCACATGAGGCTAGTGTGTTCTACTGATCAATGCCCTGACGTTG 187
OY 4394 CATTATTTGTATGAAAAAGAAAGACTGGATTAACTCTATCAAGTAGAGACATG 4453
    |||||||
DB 188 CATTATTTGTATGAAAAAGAAAGACTGGATTAACTCTATCAAGTAGAGACATG 247
OY 4454 AGACCAATGTGTCTACATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4513
    |||||||
DB 248 AGACCAATGTGTCTACATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307
OY 4514 TTTATGTGAGACAGATCTCATTTCTGTTGCTAGGAGGATGAGGAGGATCTG 4573
    |||||||
DB 308 TTTATGTGAGACAGATCTCATTTCTGTTGCTAGGAGGATGAGGAGGATCTG 367
OY 4574 GCTCACTGCACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4633
    |||||||
DB 368 GCTCACTGCACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427
OY 4634 GCTGGGATCACTGGCACAACACCATGCTGCTGCTGCTGCTGCTGCTGCTG 4693
    |||||||
DB 428 GCTGGGATCACTGGCACAACACCATGCTGCTGCTGCTGCTGCTGCTGCTG 487
OY 4694 GCTTTCACCATGTTGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4753
    |||||||
DB 488 GCTTTCACCATGTTGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
OY 4754 CGGCTCCCAAAAGTCTGGGATTTACAGATGTGAGCCAGCCAGCCAGCCAGCC 4813
    |||||||
DB 548 CGGCTCCCAAAAGTCTGGGATTTACAGATGTGAGCCAGCCAGCCAGCCAGCC 607
OY 4814 TTTATA-CCAATTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4869
    |||||||
DB 608 TTTATACCAATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667
OY 4870 GATCTGGAAGGAGGATTAATAGTATAGCTGCTGCTGCTGCTGCTGCTGCTG 4929
    |||||||
DB 668 GATCTGGAAGGAGGATTAATAGTATAGCTGCTGCTGCTGCTGCTGCTGCTG 727
OY 4930 TCGGCAAAAACCTCTCTCAAAAATAATGACTTCAATTTGATTCATGAATTCAC 4989
    |||||||
DB 728 TCGGCAAAAACCTCTCTCTCAAAAATAATGACTTCAATTTGATTCATGAAT 787

RESULT 13
B0018619/c 712 bp mRNA linear EST 27-MAR-2002
LOCUS B0018619 IMAGE:5823821.3, mRNA sequence.
DEFINITION B0018619.1 GI:19753896
ACCESSION B0018619
VERSION B0018619.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 712)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA-yes.

```

```

FEATURES
    source
        location/Qualifiers
            1..712
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5823821"
            /clone.lib="NCI-CCAP.DH1"
            /tissue.type="Metastatic Chondrosarcoma"
            /dev_stage="Adult"
            /lab_host="DH10B (Life Technologies)"
            /note="Organ: Lung; Vector: pTR73-Pac (Pharmacia) with a
            modified polylinker; Site_1: EcoR I; Site_2: Not I;
            NCI-CCAP.DH1 is a normalized cDNA library containing the
            following tissue(s): VS-8 Cell line from Metastatic
            Chondrosarcoma in Lung. The library was constructed
            according to Bonaldo, Lennon and Soares, Genome Research,
            6:791-806, 1996. First strand cDNA synthesis was primed
            with an oligo-dT primer containing a Not I site. Double
            stranded cDNA was ligated to an EcoR I adaptor, digested
            with Not I, and cloned directionally into pTR73-Pac
            vector. The oligonucleotide used to prime the synthesis of
            first strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dt)18 tail. The
            sequence tag for this library is AGATCATTCG.
            TAG_LIB=UI-H-DH1
            TAG_TISSUE=Lung
            TAG_SEQ=AGATCATTCG"
BASE COUNT 174 a 163 c 192 g 181 t 2 others
ORIGIN
Query Match 11.1%; Score 687; DB 14; Length 712;
Best Local Similarity 99.0%; Pred. No. 1.3e-85;
Matches 690; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
OY 5226 GCTTCCTTCCCTCACTGGAAGAGCCCTTATTTGATTTCACTGTGTGAGCCCTAGCCCTTC 5285
    |||||||
DB 699 GCTTCCTTCCCTCACTGGAAGAGCCCTTATTTGATTTCACTGTGTGAGCCCTAGCCCTTC 640
OY 5286 ATTCTCGAATTCCTCCCACTCCAGCCCTTCCCAAGAGAGACTAGTGCCTCGCATTC 5345
    |||||||
DB 639 ATTCTCGAATTCCTCCCACTCCAGCCCTTCCCAAGAGAGACTAGTGCCTCGCATTC 580
OY 5346 ACCAAGATGGGATTTGGCTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5405
    |||||||
DB 579 ACCAAGATGGGATTTGGCTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 520
OY 5406 TTGCTCTGAAGACACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5465
    |||||||
DB 519 TTGCTCTGAAGACACACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 460
OY 5466 AAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5525
    |||||||
DB 459 AAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400
OY 5526 CTGAGGACCTCGGCTCTTTCTTTTAAAGCCGAGCCCACTTACTATAAATTCAGG 5585
    |||||||
DB 399 CTGAGGACCTCGGCTCTTTCTTTTAAAGCCGAGCCCACTTACTATAAATTCAGG 340
OY 5586 GTTCACTGGAAGAGTAAGTGCATTTGTTGAAGCCTACTGATGCTGACCCACTGCTCA 5645
    |||||||
DB 339 GTTCACTGGAAGAGTAAGTGCATTTGTTGAAGCCTACTGATGCTGACCCACTGCTCA 280
OY 5646 TCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5705
    |||||||
DB 279 TCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 220
OY 5706 TGTGATTCATTCACAGAGGAAAGGCTCAAGGAGAGTCAACTGGAGCAAGCAACAG 5765
    |||||||
DB 219 TGTGATTCATTCACAGAGGAAAGGCTCAAGGAGAGTCAACTGGAGCAAGCAACAG 160
OY 5766 CCACCGGACATGGCTTGTGTAAGGTTAGCAGACTGCTGTGTGATCTGCAGTGCCTTC 5825
    |||||||
DB 159 CCACCGGACATGGCTTGTGTAAGGTTAGCAGACTGCTGTGTGATCTGCAGTGCCTTC 100

```

QY 5826 ACTGGAATATTTATTCATGACAGATCTTTAGTGCATTTATTCATTTCTGTG 5885
 |||||||
 Db 99 ACTGAAATATTTATTCATGACAGATCTTTAGTGCATTTATTCATTTCTGTG 40
 |||||||
 QY 5886 CTTTAATAAACAATGTACCAAAAACAGTATCA 5922
 |||||||
 Db 39 CTTTAATAAACAATGTACCAAAAACAGTATCA 3
 |||||||
 RESULT 14
 Bg118850 835 bp mRNA linear EST 30-JAN-2001
 LOCUS 602347846F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443010 5',
 mRNA sequence.
 ACCESSION Bg118850
 VERSION Bg118850.1 GI:12612356
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 835)
 NIH-MGC http://imgc.ncl.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10216 row: k column: 11
 High quality sequence stop: 717.
 Location/Qualifiers
 1. 835
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4443010"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10b (phage-resistant)"
 /note="Organ: Liver; Vector: PCMV-SPOrt6; Site_1: NCI;
 Site_2: Sall; Cloned unidirectionally; oligo-dt primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."
 BASE COUNT 182 a 236 c 209 g 207 t 1 others
 ORIGIN
 Query Match 11.1%; Score 683.6; DB 12; Length 835;
 Best Local Similarity 95.0%; Pred. No. 3.4e-85;
 Matches 783; Conservative 0; Mismatches 29; Indels 12; Gaps 7;
 QY 5007 ACTTGTTCACATGCAATTAATGAAGCAATTTATAGTCTCTTAATGGCGTGTACTGCA 5066
 |||||||
 Db 1 ACTTGTTCACATGCAATTAATGAAGCAATTTATAGTCTCTTAATGGCGTGTACTGCA 60
 |||||||
 QY 5067 AGACCTCTTGAACACTTCCAGAGATAGATATTTAAGTCATGCCCTTGCGCTTA 5126
 |||||||
 Db 61 AGACCTCTTGAACACTTCCAGAGATAGATATTTAAGTCATGCCCTTGCGCTTA 120
 |||||||
 QY 5127 TGACACCTTCCCTTCGAAAGTCTGGTTCCTGCCAGTGACCTTGCGCTTGAGCG 5186
 |||||||
 Db 121 TGACACCTTCCCTTCGAAAGTCTGGTTCCTGCCAGTGACCTTGCGCTTGAGCG 180
 |||||||
 QY 5187 AGATGTGACCCCTGCATTAAGGCGCAAGAGGCGCTGCGGCTTCCTCCCTCACTGAGA 5246
 |||||||
 Db 181 AGATGTGACCCCTGCATTAAGGCGCAAGAGGCGCTGCGGCTTCCTCCCTCACTGAGA 240
 |||||||
 QY 5247 GCCCTATTGTAATTCAGTGTGAGGCCCTAGCCCTCATTCGTGACATTTCCCAACCT 5306
 |||||||

|||||||
 Db 241 GCCCTATTGTAATTCAGTGTGAGGCCCTAGCCCTCATTCGTGACATTTCCCAACCT 300
 |||||||
 QY 5307 CCCAGCCCTTCCAGAGAGCTAGTGGCCCTGATTCACCCCAAGTGGGATGGGCTT 5366
 |||||||
 Db 301 CCCAGCCCTTCCAGAGAGCTAGTGGCCCTGATTCACCCCAAGTGGGATGGGCTT 359
 |||||||
 QY 5367 CTTTAGGCTGCTACTTGTGCACATCCAGCATCACTGTTCCTGCAAGACACAGT 5426
 |||||||
 Db 360 CTTTAGGCTGCTACTTGTGCACATCCAGCATCACTGTTCCTGCAAGACACAGT 419
 |||||||
 QY 5427 GGGCATTTTCTTCATGAGGAGGCTCAAACTCTGGACAAAGTTGCTGCTCTGAGACC 5486
 |||||||
 Db 420 GGGCATTTTCTTCATGAGGAGGCTCAAACTCTGGACAAAGTTGCTGCTCTGAGACC 479
 |||||||
 QY 5487 AGTATTCTCGAGAGCTGTGCTCAGTGAAGGGGCCAGCTGAGAGAACCTTGCTCTTT 5546
 |||||||
 Db 480 AGTATTCTCGAGAGCTGTGCTCAGTGAAGGGGCCAGCTGAGAGAACCTTGCTCTTT 539
 |||||||
 QY 5547 CTTTAAAGCCAGGCCCCACTTACATTAACATTTTCAGGGTCACTGGAACAGTG- AAGT 5605
 |||||||
 Db 540 CTTTAAAGCCAGGCCCCACTTACATTAACATTTTCAGGGTCACTGGAACAGTG- AAGT 559
 |||||||
 QY 5606 GCCATTGTGTT- GAAGCTTCTGATGCTCCAGCCCACT- GCTCATCCACGTGCTT- GCCA 5661
 |||||||
 Db 600 GCCATTGTGTTGAAGCCTTACTGATGCTCCAGCCCACTGCTCATCCAGCTGCTTCCCA 659
 |||||||
 QY 5662 TGCCCTACGAGAAAGCCAGGCCAGTCAAGGAC- TGCTCTTAATGCTGTGCTCATTTGACA 5720
 |||||||
 Db 660 TGCCCTACGAGAAAGCCAGGCCAGTCAAGGAC- TGCTCTTAATGCTGTGCTCATTTGACA 719
 |||||||
 QY 5721 GAAGGGAAGGCTCAAGGAAGTCACTGGGACAGCA- ----CAAGCCACCGGACA 5775
 |||||||
 Db 720 GAAGGGAAGGCTCAAGGAAGTCACTGGGACAGCA- ----CAAGCCACCGGACA 779
 |||||||
 QY 5776 TGCCCTTGTGAAGGTTAGCAGACTGCTGTGCTGTGATCTGCA 5819
 |||||||
 Db 780 TGCCCTTGTGAAGGTTAGCAGACTGCTGTGCTGTGATCTGCA 823
 |||||||
 RESULT 15
 B1457122 683 bp mRNA linear EST 21-AUG-2001
 LOCUS B1457122
 DEFINITION 603185367F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5258105 5',
 mRNA sequence.
 ACCESSION B1457122
 VERSION B1457122.1 GI:15247778
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 683)
 NIH-MGC http://imgc.ncl.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM1869 row: e column: 18
 High quality sequence stop: 678.
 Location/Qualifiers
 1. 683
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5258105"
 /clone_lib="NIH_MGC_42"

/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(C). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library. |"

BASE COUNT 170 a 205 c 171 g 137 t
ORIGIN

Query Match 11.0%: Score 678.8: DB 13: Length 683:
Best Local Similarity 99.7%: Pred.No.1.8e-84;
Matches 680: Conservative 0: Mismatches 2: Indels 0: Gaps 0:
QY 1103 GTGAGAGAGAGAGAGAGCGGTTGAATACATCCGGGCTCCACACCAACCCGAGG 1162
DB 2 GTGAGAGAGAGAGAGAGCGGTTGAATACATCCGGGCTCCACACCAACCCGAGG 61
QY 1163 TGTTCAGCTGAGAGACAAGACCTGGGACATGGGGGAACCTTCAACTCTCTCTGC 1222
DB 62 TGTTCAGCTGAGAGACAAGACCTGGGACATGGGGGAACCTTCAACTCTCTCTGC 121
QY 1223 AAGCGTGTACCAAGATGCCAAAGTCCAGGATCTCCGGCTGCAGTTCCAAATTGTTGG 1282
DB 122 AAGCGTGTACCAAGATGCCAAAGTCCAGGATCTCCGGCTGCAGTTCCAAATTGTTGG 181
QY 1283 TCCAACTCCACAAATGAAGCAATTAATCTACGTGGTGGTACTGAGTAATGAGCGAG 1342
DB 182 TCCAACTCCACAAATGAAGCAATTAATCTACGTGGTGGTACTGAGTAATGAGCGAG 241
QY 1343 CCATGTCACTCACTCGACGACGCGCCGTCAAACAGACCGCAAGTTGTCCCTGGCT 1402
DB 242 CCATGTCACTCACTCGACGACGCGCCGTCAAACAGACCGCAAGTTGTCCCTGGCT 301
QY 1403 GTTTCGTGTGTAGAAATTCGGACCTGCAAGTACCACTCACTGACATCTGGCTCA 1462
DB 302 GTTTCGTGTGTAGAAATTCGGACCTGCAAGTACCACTCACTGACATCTGGCTCA 361
QY 1463 AACACAAATCTCTCTCTTGTGATGATCTGACAGCTGCTGGATGATGGAAAAA 1522
DB 362 AACACAAATCTCTCTCTTGTGATGATCTGACAGCTGCTGGATGATGGAAAAA 421
QY 1523 CCATAAGCTGCACAGACACCGGTACTGCAAAAGAAATCTACTACTCCAGTGCACA 1582
DB 422 CCATAAGCTGCACAGACACCGGTACTGCAAAAGAAATCTACTACTCCAGTGCACA 481
QY 1583 GTGACATCTCTCACTGCTGTGGAGCTGATGACTTCTCTGGAAAGCTGTGGTCCCA 1642
DB 482 GTGACATCTCTCACTGCTGTGGAGCTGATGACTTCTCTGGAAAGCTGTGGTCCCA 541
QY 1643 AGGACAGGCTCAGCCTGTGTGTGTCAGCCAGAAAGCTGCAGCAGCATACACAGAGA 1702
DB 542 AGGACAGGCTCAGCCTGTGTGTGTCAGCCAGAAAGCTGCAGCAGCATACACAGAGA 601
QY 1703 AGCCTGCACACAGCTTCACTACCTGTGGCCAGTGCATATCCACAGAGACTGT 1762
DB 602 AGCCTGCACACAGCTTCACTACCTGTGGCCAGTGCATATCCACAGAGACTGT 661
QY 1763 ACTTGGCTCTCTTCTGCCCCGG 1784
DB 662 ACTTGGCTCTCTTCTGCCCCGG 683

Search completed: November 11, 2002, 11:24:42
Job time : 7744 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:32:51 : Search time 77 Seconds

(without alignments)
2237.083 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394

Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKDPDPLNTQEPMEPAE 836

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTOZOA:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIA:*
17: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4394	100.0	836	4	Q96Q07
2	4385	99.8	836	4	Q95V8
3	3411	77.6	649	4	Q9H8C2
4	2047	46.6	392	4	Q9H676
5	1779	40.5	343	4	Q8WU91
6	1594.5	36.3	384	11	Q921M9
7	140	3.2	746	5	001654
8	137.5	3.1	719	13	Q9PVY2
9	137.5	3.1	3623	11	Q9244
10	127	2.9	832	11	Q9JLR4
11	126	2.9	595	4	Q96P01
12	123.5	2.8	986	5	Q76355
13	123.5	2.8	3620	6	Q9T053
14	123	2.8	724	10	Q93VB7
15	118.5	2.7	1406	4	Q9Y216
16	118	2.7	2766	11	Q9QZ88

17	118	2.7	3564	11	Q923L3	Q923L3 mus musculus
18	118	2.7	3623	4	Q60494	Q60494 homo sapien
19	117.5	2.7	774	5	Q9VAA2	Q9VAA2 drosophila
20	117.5	2.7	1464	5	Q24132	Q24132 drosophila
21	117.5	2.7	1464	5	Q23995	Q23995 drosophila
22	117.5	2.7	1464	5	Q9VC47	Q9VC47 drosophila
23	117	2.7	597	6	Q95LH0	Q95LH0 pan troglod
24	117	2.7	603	5	Q22075	Q22075 caenorhabd
25	117	2.7	1809	5	Q9ALB8	Q9ALB8 caenorhabd
26	116.5	2.7	1027	5	Q17657	Q17657 caenorhabd
27	116.5	2.7	1424	10	Q1LTR5	Q1LTR5 arbidops
28	116.5	2.7	3121	13	Q42269	Q42269 brachydant
29	116.5	2.7	4215	5	Q9W332	Q9W332 drosophila
30	115.5	2.6	407	11	Q9D224	Q9D224 mus musculu
31	115.5	2.6	1210	5	Q8T1W5	Q8T1W5 dictyostell
32	115.5	2.6	1296	11	Q91VX3	Q91VX3 mus musculu
33	114.5	2.6	701	13	Q90X05	Q90X05 xenopus lae
34	114.5	2.6	1196	13	Q98TF1	Q98TF1 cyprinus ca
35	114.5	2.6	1573	10	Q94LS7	Q94LS7 oryza sativ
36	114.5	2.6	3898	12	Q11994	Q11994 muscosal dis
37	114	2.6	449	5	Q95RA3	Q95RA3 drosophila
38	114	2.6	16215	5	Q9NFS3	Q9NFS3 drosophila
39	113.5	2.6	688	13	Q9PVY4	Q9PVY4 xenopus lae
40	113.5	2.6	838	5	Q9VOA9	Q9VOA9 drosophila
41	113	2.6	1224	3	Q13637	Q13637 schizosacch
42	113	2.6	3139	11	Q91ZHO	Q91ZHO mus musculu
43	113	2.6	3129	11	Q8VH00	Q8VH00 mus musculu
44	112.5	2.6	701	13	Q8UVQ4	Q8UVQ4 xenopus lae
45	112.5	2.6	1018	10	Q9LSX6	Q9LSX6 arbidops

ALIGNMENTS

RESULT 1

Q96Q07 PRELIMINARY: PRT: 836 AA.

AC Q96Q07: 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
ID CUB domain containing protein 1.
GN CDCP1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21359660; PubMed-1146621;
RA Scherl-Mostaguer M., Sommergruber W., Abseher R., Hauptmann R.,
RA Ambros P., Schweifer N.;
RT *Identification of a novel gene, CDCP1, overexpressed in human
RT colorectal cancer.*
RL Oncogene 20:4402-4408 (2001).
DR EMBL: AY026461; AAK02058.1;
SQ SEQUENCE 836 AA; 92873 MW; FB4D2DBBD35C519 CRC64;

Query Match 100.0%; Score 4394; DB 4; Length 836;
Best local similarity 100.0%; Pred. No. 0;
Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLNCVSTALLGVLLGAARLPKGAEPFETALPRESNITVLIKIGTPTLLAKPCYIYI 60
DB 1 MAGLNCVSTALLGVLLGAARLPKGAEPFETALPRESNITVLIKIGTPTLLAKPCYIYI 60
QY 61 SRKHTMTSIKSGERTVFVFSCQSPENHVFIEIKKIDMSGCPGEGVQLDPSTSLPT 120
DB 61 SRKHTMTSIKSGERTVFVFSCQSPENHVFIEIKKIDMSGCPGEGVQLDPSTSLPT 120
QY 121 LNRFTFMDVKAHKSIGLEQFSIPRLQIGPSCPDGVTHSISGRIDATVVRIGFCFN 180
DB 121 LNRFTFMDVKAHKSIGLEQFSIPRLQIGPSCPDGVTHSISGRIDATVVRIGFCFN 180

```

QY 181 GTVSRIKMOGVKMAHLHPFHPHNRNVSQFSIANRSSIKRLCIIESVFEGBSATLMSANY 240
DB 181 GTVSRIKMOGVKMAHLHPFHPHNRNVSQFSIANRSSIKRLCIIESVFEGBSATLMSANY 240
QY 241 PEQFPDELTMTQFPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEDK 300
DB 241 PEQFPDELTMTQFPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEDK 300
QY 301 QPNNMAGNFNLISQGCCDQDQSPGILRLQFOVLVQHPQNSNKIYYVDLSNERAMSLTIE 360
DB 301 QPNNMAGNFNLISQGCCDQDQSPGILRLQFOVLVQHPQNSNKIYYVDLSNERAMSLTIE 360
QY 361 PRPVKOSRKFPVPCFCVLESRTCSNLTTLTSSGSKHKSIFLCDDLTRLMNVKERTISCTDH 420
DB 361 PRPVKOSRKFPVPCFCVLESRTCSNLTTLTSSGSKHKSIFLCDDLTRLMNVKERTISCTDH 420
QY 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAQKLOQHTHEKPCNTSF 480
DB 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAQKLOQHTHEKPCNTSF 480
QY 481 SYLVASAIIPSDILYFGSFCGSGSIKQIQVKNISVTLTRFAPSFOEASRQGLTVSFIPY 540
DB 481 SYLVASAIIPSDILYFGSFCGSGSIKQIQVKNISVTLTRFAPSFOEASRQGLTVSFIPY 540
QY 541 FKREGVFTVTPDKSKYYLTPTMMDRGLPSTLTVSNMISVPRDQVACLTFFEKERSGVYQC 600
DB 541 FKREGVFTVTPDKSKYYLTPTMMDRGLPSTLTVSNMISVPRDQVACLTFFEKERSGVYQC 600
QY 601 TGRAFMIOQORTRAEIEFLDEVDLPKPSFHHHSFVNINSNCSPTSGKOLDLFSVTLL 660
DB 601 TGRAFMIOQORTRAEIEFLDEVDLPKPSFHHHSFVNINSNCSPTSGKOLDLFSVTLL 660
QY 661 PRVVDLTVILIAAVGGVLLLSALGLIICCVKKKKTKNGPVGIIYNGINTEMPPROPK 720
DB 661 PRVVDLTVILIAAVGGVLLLSALGLIICCVKKKKTKNGPVGIIYNGINTEMPPROPK 720
QY 721 KFOGRKNDNSHYAVIETMYGHLQDSSGSFLOEVDYTRPFQGTWGCVPSPPTIC 780
DB 721 KFOGRKNDNSHYAVIETMYGHLQDSSGSFLOEVDYTRPFQGTWGCVPSPPTIC 780
QY 781 SRAPTAKLATEEPPRPSPSESEPTFFSHPNNGDVSSKTDIPLLTQPMPEPAE 836
DB 781 SRAPTAKLATEEPPRPSPSESEPTFFSHPNNGDVSSKTDIPLLTQPMPEPAE 836

RESULT 2
Q9H5V8 PRELIMINARY; PRT; 836 AA.
AC Q9H5V8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA: FLJ22969 fis, clone KAT10759.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Odayashi M., Nishitani T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono M., Sugano S.,
RA "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK026622; BAB1511.1; - 9B980475C3E5C4C8 CRG4;
SQ SEQUENCE 836 AA; 92874 MW; 9B980475C3E5C4C8 CRG4;

Query Match 99.8%; Score 4385; DB 4; Length 836;
Best local Similarity 99.8%; Pred. No. 0;
Matches 834; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGLNCGVSTALIGVLLGAAARLPRGAFAEIALPRESNTTVLIKGTPTLLAKPCYIYI 60

```

```

DB 1 MAGLNCGVSTALIGVLLGAAARLPRGAFAEIALPRESNTTVLIKGTPTLLAKPCYIYI 60
QY 61 SKRHITMLSIKSGERIVTFPSQSPENHVFIELOKNIDCSGCPGCEVOLQSTSLPT 120
DB 61 SKRHITMLSIKSGERIVTFPSQSPENHVFIELOKNIDCSGCPGCEVOLQSTSLPT 120
QY 121 LNRTFIMDVKAHKSIGLELOFISIPRLQIGPGSCPDGVTHSISGRIDATVVRIGTFCSN 180
DB 121 LNRTFIMDVKAHKSIGLELOFISIPRLQIGPGSCPDGVTHSISGRIDATVVRIGTFCSN 180
QY 181 GTVSRIKMOGVKMAHLHPFHPHNRNVSQFSIANRSSIKRLCIIESVFEGBSATLMSANY 240
DB 181 GTVSRIKMOGVKMAHLHPFHPHNRNVSQFSIANRSSIKRLCIIESVFEGBSATLMSANY 240
QY 241 PEQFPDELTMTQFPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEDK 300
DB 241 PEQFPDELTMTQFPVPAHLRASVSFLNFMNSCERKEEVEYYIGSTINPEVFKLEDK 300
QY 301 QPNNMAGNFNLISQGCCDQDQSPGILRLQFOVLVQHPQNSNKIYYVDLSNERAMSLTIE 360
DB 301 QPNNMAGNFNLISQGCCDQDQSPGILRLQFOVLVQHPQNSNKIYYVDLSNERAMSLTIE 360
QY 361 PRPVKOSRKFPVPCFCVLESRTCSNLTTLTSSGSKHKSIFLCDDLTRLMNVKERTISCTDH 420
DB 361 PRPVKOSRKFPVPCFCVLESRTCSNLTTLTSSGSKHKSIFLCDDLTRLMNVKERTISCTDH 420
QY 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAQKLOQHTHEKPCNTSF 480
DB 421 RYCORSYSIQVPSDILHLPELHDEFSWKLLVPRKDRSLVLVPAQKLOQHTHEKPCNTSF 480
QY 481 SYLVASAIIPSDILYFGSFCGSGSIKQIQVKNISVTLTRFAPSFOEASRQGLTVSFIPY 540
DB 481 SYLVASAIIPSDILYFGSFCGSGSIKQIQVKNISVTLTRFAPSFOEASRQGLTVSFIPY 540
QY 541 FKREGVFTVTPDKSKYYLTPTMMDRGLPSTLTVSNMISVPRDQVACLTFFEKERSGVYQC 600
DB 541 FKREGVFTVTPDKSKYYLTPTMMDRGLPSTLTVSNMISVPRDQVACLTFFEKERSGVYQC 600
QY 601 TGRAFMIOQORTRAEIEFLDEVDLPKPSFHHHSFVNINSNCSPTSGKOLDLFSVTLL 660
DB 601 TGRAFMIOQORTRAEIEFLDEVDLPKPSFHHHSFVNINSNCSPTSGKOLDLFSVTLL 660
QY 661 PRVVDLTVILIAAVGGVLLLSALGLIICCVKKKKTKNGPVGIIYNGINTEMPPROPK 720
DB 661 PRVVDLTVILIAAVGGVLLLSALGLIICCVKKKKTKNGPVGIIYNGINTEMPPROPK 720
QY 721 KFOGRKNDNSHYAVIETMYGHLQDSSGSFLOEVDYTRPFQGTWGCVPSPPTIC 780
DB 721 KFOGRKNDNSHYAVIETMYGHLQDSSGSFLOEVDYTRPFQGTWGCVPSPPTIC 780
QY 781 SRAPTAKLATEEPPRPSPSESEPTFFSHPNNGDVSSKTDIPLLTQPMPEPAE 836
DB 781 SRAPTAKLATEEPPRPSPSESEPTFFSHPNNGDVSSKTDIPLLTQPMPEPAE 836

RESULT 3
Q9H8C2 PRELIMINARY; PRT; 649 AA.
AC Q9H8C2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA FLJ13772 fis, clone PLACE4000300.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-PLACENTA;
RA Isono M., Sugano S., Takahashi-Fujii A., Hara H.,
RA Nishikawa T., Nagai K., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

```

RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Makamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK023834; BAB1695.1; -
 FT NON_TER 392
 SO SEQUENCE 649 AA; 72579 MW; 0F4404AD31D60E9A CRC64;

Query Match 77.6%; Score 3411; DB 4; Length 649;
 Best Local Similarity 99.5%; Pred. No. 1,3e-293;
 Matches 646; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 188 MCGVGMALHLPFHPHNSGFSIANSRSIKRCLITISVEGEGSATLMSANYPEGEPD 247
 |||||
 DB 1 MCGVGMALHLPFHPHNSGFSIANSRSIKRCLITISVEGEGSATLMSANYPEGEPD 60
 QY 248 ELMTQFVVPVPAHLRASVSLFNLSNCEKREVEYYIPGSTNPEVFKLEDKOPGMAG 307
 |||||
 DB 61 ELMTQFVVPVPAHLRASVSLFNLSNCEKREVEYYIPGSTNPEVFKLEDKOPGMAG 120
 QY 308 NEFLSLGGCDODASPGILRLQFOVLVOHPONESNKIYYVDLSNERAMSLTIEPRPV 367
 |||||
 DB 121 NEFLSLGGCDODASPGILRLQFOVLVOHPONESNKIYYVDLSNERAMSLTIEPRPV 180
 QY 368 RKEVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 427
 |||||
 DB 181 RKEVPCFVCELESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 240
 QY 428 YSLQVPSDILHLPVELHDFSWKLLVPRKRLSLVLPKQRLQOHTHEKPCMTSFSYL 487
 |||||
 DB 241 YSLQVPSDILHLPVELHDFSWKLLVPRKRLSLVLPKQRLQOHTHEKPCMTSFSYL 300
 QY 488 IPSODILFSGFCGSGSIKOIOVKONISVTLRTAPSPQOASRQGLTVSFIPFKEGV 547
 |||||
 DB 301 IPSODILFSGFCGSGSIKOIOVKONISVTLRTAPSPQOASRQGLTVSFIPFKEGV 360
 QY 548 TVTPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKESGVVCCOTGRA 607
 |||||
 DB 361 TVTPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKESGVVCCOTGRA 420
 QY 608 IOEORTAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKLDLFSVTLPRVDT 667
 |||||
 DB 421 IOEORTAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKLDLFSVTLPRVDT 480
 QY 668 VILIAVGGVLLSLGLITICVKKKKTKNGPRAVGININTEMPQPKFQGRK 727
 |||||
 DB 481 VILIAVGGVLLSLGLITICVKKKKTKNGPRAVGININTEMPQPKFQGRK 540
 QY 728 DNDSHYAVIEDTMYGHLLODSSGFLQPEVDTYRPFQGTMGVCPSPPTICSRAPTAK 787
 |||||
 DB 541 DNDSHYAVIEDTMYGHLLODSSGFLQPEVDTYRPFQGTMGVCPSPPTICSRAPTAK 600
 QY 788 LATEPPPSPESESEPTFSHPNNGDVSSKDTDIPLNTQDPMBAE 836
 |||||
 DB 601 LATEPPPSPESESEPTFSHPNNGDVSSKDTDIPLNTQDPMBAE 649

RESULT 4

Q9H676 PRELIMINARY; PRT; 392 AA.

AC Q9H676
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 CDNA: FLJ22534 f1s, clone HRC13020 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK026187; BAB1538.1; -
 FT NON_TER 392
 SO SEQUENCE 392 AA; 44031 MW; 3E21C8B2703E52F3 CRC64;

Query Match 46.6%; Score 2047; DB 4; Length 392;
 Best Local Similarity 99.7%; Pred. No. 6.4e-173;
 Matches 391; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 305 MAGFNLSLGGCDODASPGILRLQFOVLVOHPONESNKIYYVDLSNERAMSLTIEPRV 364
 |||||
 DB 1 MAGFNLSLGGCDODASPGILRLQFOVLVOHPONESNKIYYVDLSNERAMSLTIEPRV 60
 QY 365 KQSRKFFVGCFCLESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 424
 |||||
 DB 61 KQSRKFFVGCFCLESRTCSNLTLSGSKHKISFLCDDLRLMMNVEKTICTDHRVCO 120
 QY 425 RKYSLOVPSDILHLPVELHDFSWKLLVPRKRLSLVLPKQRLQOHTHEKPCMTSFSYL 484
 |||||
 DB 121 RKYSLOVPSDILHLPVELHDFSWKLLVPRKRLSLVLPKQRLQOHTHEKPCMTSFSYL 180
 QY 485 ASAIPSDILFSGFCGSGSIKOIOVKONISVTLRTAPSPQOASRQGLTVSFIPFKEE 544
 |||||
 DB 181 ASAIPSDILFSGFCGSGSIKOIOVKONISVTLRTAPSPQOASRQGLTVSFIPFKEE 240
 QY 545 GVEFTVPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKESGVVCCOTGRA 604
 |||||
 DB 241 GVEFTVPDTSKYVLRTPMMDRGLPSLTSVSWNISVPRDOVACLTFFKESGVVCCOTGRA 300
 QY 605 FMIIQORTAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKLDLFSVTLPRV 664
 |||||
 DB 301 FMIIQORTAEIEFSLDEVDLPKPSFHHSFWVNISNCSPTSGKLDLFSVTLPRV 360
 QY 665 DLTVILAAVGGVLLSLGLITICVKKKK 696
 |||||
 DB 361 DLTVILAAVGGVLLSLGLITICVKKKK 392

RESULT 5

O8W091 PRELIMINARY; PRT; 343 AA.

AC O8W091
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Similar to hypothetical protein FLJ22969.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-KIDNEY;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021099; AAH21099.1; -
 KW Hypothetical protein.
 SO SEQUENCE 343 AA; 37817 MW; 3E4E13379DD94D1B CRC64;

Query Match 40.5%; Score 1779; DB 4; Length 343;
 Best Local Similarity 99.4%; Pred. No. 3.1e-149;
 Matches 341; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLNGVSTALGVLIGARLPKGAFAEILPRESNTVILIKGTPTLAKPCYIV 60
 |||||
 DB 1 MAGLNGVSTALGVLIGARLPKGAFAEILPRESNTVILIKGTPTLAKPCYIV 60
 QY 61 SKRHITMLSIKSGERIVTFSCSPENHFYIEIOKNIDCSGCPGEGVQLQPSILP 60
 |||||
 DB 61 SKRHITMLSIKSGERIVTFSCSPENHFYIEIOKNIDCSGCPGEGVQLQPSILP 60

QY 121 LNRFIDVAKHKSIGLELPSIRLROIGGSCPDGVTHSISGRIDATVVRIGTRCSN 180
 DB 121 LNRFIDVAKHKSIGLELPSIRLROIGGSCPDGVTHSISGRIDATVVRIGTRCSN 180
 QY 181 GTVARIKMOGEVKAALHPHFPNNSGFSIANSRSIKRLCIISVFECSATLMSANY 240
 DB 181 GTVARIKMOGEVKAALHPHFPNNSGFSIANSRSIKRLCIISVFECSATLMSANY 240
 QY 241 PEGPEDELMTQFVPAHLRASVFLNLSNCRKEERYEYIPGSTNPEVEKLEDK 300
 DB 241 PEGPEDELMTQFVPAHLRASVFLNLSNCRKEERYEYIPGSTNPEVEKLEDK 300
 QY 301 QPGMAGNFNLSLOGCDODASPGILRLQFOVLYVQHPONESNK 343
 DB 301 QPGMAGNFNLSLOGCDODASPGILRLQFOVLYVQHPONESNK 343

RESULT 6

Q921M9 PRELIMINARY; PRT; 384 AA.
 AC Q921M9
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Unknown (Protein for IMAGE:3709337) (Fragment).
 GN AA409659.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC011340; AAH11340.1;
 DR MGD: MGI:2142933; AA409659.
 FT NON_TER 1
 SQ SEQUENCE 384 AA; 42406 MW; 4B5F17D30FE16080 CRC64;

Query Match 36.3%; Score 1594.5; DB 11; Length 384;
 Best Local Similarity 77.8%; Pred. No. 8.8e-133;
 Matches 301; Conservative 33; Mismatches 48; Indels 3; Gaps 3;

QY 450 LLVPRKRLSLVPAQKIQHTHEKPCNTSFSYVASAIPSDILYFGSFCPSGSIKQIQV 509
 DB 1 LLVPRKRLSLVPAQKIQHTHEKPCNTSFSYVASAIPSDILYFGSFCPSGSIKQIQV 60
 QY 510 KQNTSVLRFAPSPGOEASROGLTVSFIPEKKEGVTVTPDMKSKVYLTPMMDGLP 569
 DB 61 KQNTSVLRFAPSPGOEASROGLTVSFIPEKKEGVTVTPDMKSKVYLTPMMDGLP 120
 QY 570 SLTSVSNINISVPDQVACTFEKERSGVVQGTGRAFMIIQORTRAEIEFSLDSDVLPK 629
 DB 121 ALSSVSNINISVPDQVACTFEKERSGVVQGTGRAFMIIQORTRAEIEFSLDSDVLPK 180
 QY 630 SFHHSEFVNINISCSPTSGKQDLDFSVTLTPRTVDTLVIIAAVGGVLLLSALGLIC 689
 DB 181 SFHHSEFVNINISCSPTSGKQDLDFSVTLTPRTVDTLVIIAAVGGVLLLSALGLIC 239
 QY 690 CVKRRKKTKNGRAVGIYNGININTEMPQPKKQKGRKDNDSHYAYIEDIMVGHLLD 749
 DB 240 FV-KRRKKTKNGRAVGIYNGININTEMPQPKKQKGRKDNDSHYAYIEDIMVGHLLD 297
 QY 750 SSGSFLQPEVNTYRPGGTGVCPPSPPTICSRAPTKAKLAEPPSPSPSESEPTYS 809
 DB 298 SSGSFLQPEVNTYRPGGTGVCPPSPPTICSRAPTKAKLAEPPSPSPSESEPTYS 357
 QY 810 HPNNGDVSSKDTDIPLANTOEPMEPAE 836
 DB 358 HPNNGDVSSKDTDIPLANTOEPMEPAE 384

RESULT 7

001654
 AC 001654
 ID 001654
 DT 01-JUL-1997 (TREMblrel. 04, Created)
 DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
 DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE AsnAspa.
 OS Halocynthia roretzi (Sea squirt).
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea,
 OC Stolidobranchia; Pyrosidae; Halocynthia.
 OX NCBI_TaxID=7729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEPATOPANCREAS;
 RX MEDLINE=9732374; PubMed=9177219;
 RA Uchiyama K., Sasaki M., Nonaka M.;
 RT "Ancient origin of the complement lectin pathway revealed by molecular
 cloning of mannan binding protein-associated serine protease from a
 urochordate, the Japanese Ascidian, Halocynthia roretzi.";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:6340-6345(1997).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.-ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.

DR EMBL: D88204; BAA19762.1; -.
 DR HSP; P00763; IDPO.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB domain.
 DR InterPro: IPR001881; EGF Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00084; sush1; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00134; TRYPSIN_DOM; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR EGF-like domain; Glycoprotein; Hydrolase; Hydroxylation;
 KW Serine protease.
 SQ SEQUENCE 746 AA; 84127 MW; 32B075317CF13E1 CRC64;

Query Match 3.28%; Score 140; DB 5; Length 746;
 Best Local Similarity 19.3%; Pred. No. 0.0029;
 Matches 102; Conservative 73; Mismatches 178; Indels 176; Gaps 25;

QY 220 LCIESEFEGSGATLMSA-----NYPEGPEDELMTQFVPAHLRASVFLNLS 271
 DB 11 LLYCNSVF-----SAELLTAHFGNFPSPRYPSYDINSULTININVOGGRMSIRFTFDL 66
 QY 272 -----SNCRKEERYEYIPGSTNPEVEKLEDKQPGMAGNFNLSLOGCDODASP 323
 DB 67 EDSYEDIGSC-----VYDVEITESNKTYAK-----FCGNYDL-----PFLDAPNP 108
 QY 324 GILRLQVLYVQHPONESNKTYVDLS-----NERAMSLTIERRPYKQSKK----- 369
 DB 109 -----SKFLYTQNEVRATVFSDYSISLSGFOAHQAQIDINCELMETKRTIIEEDMD 161
 QY 370 -----FVPCFCFLSERTCCSNLTLTSGSKHISFLCDDLRLMMVEXTISTCD 419
 DB 162 ELVVCCHYCRNVPGSYIC-----GCRPKFTLDANRHTCVASFCE-----OVLTDON 208
 QY 420 HRYCQKRSYSLQVSDILHLPVELHDFSWKLLVPRKD-RLSLVPAQKIQHTHEKPCNT 478
 DB 209 SGHISSEFP-----ELYAKLTDCSWITQLREGLSVNLIFRAGIEHEE-----GC 257
 QY 479 SFSYLVASAIPSDILYFGSFCPSG-----SIKQIQVKNINISVTLTFAFSPFOQ----- 526

```

Db 258 YDRLRLVLTSTDIYCGNQAPGNGTVMPTNTLVQLKFTHTDLSEVKGFSTVYSTRK 317
OY 527 -----EASHQGLVSTF---PYKEEGVFTVPDTSKYYLTPMNDRLPSLTSV----S 575
Db 318 CHALHDPNNG-SLSESHRSRYHEFDVATFSCDRGFIDII-----GVPRKLCISDS 368
OY 576 WNI5VPRDGV-AC-----LTFEKRSGVVCQ-----TGAFMTI 608
Db 369 WEHSAPICQIKSCGVQFLLDLPNSHIVRENSKITYSEVLDYTCNQWGMTSGASKWIC 428
OY 609 QEQRTAEETFSLDEVDLPKPSFHHSFVNINISN-----CSPTSGKOL 651
Db 429 ENSKIMTE-----HGLGLVAINNFNNKPVCKPICGKTL 460

```

RESULT 8

```

O9PVY2 PRELIMINARY; PRT; 719 AA.
ID O9PVY2;
AC O9PVY2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Manose-binding lectin-associated serine protease.
OS Triakis scyllium (Leopard shark) (Triakis scyllia).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes; Triakidae;
OC Triakis.
OX NCBI_Taxid=30494;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=99008558; PubMed=97944427;
RA Endo Y., Takemashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
RA Nonaka M., Fujita T.;
RT "Two lineages of manose-binding lectin-associated serine protease
RT (MASP) in vertebrates.";
RL J. Immunol. 161:4924-4930(1998).
CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRIPSLIN FAMILY.
DR EMBL: AB009074; BA06867.1; -.
DR HSSP: P00736; IAPQ.
DR MEROPS: S01.132; -.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00084; Sushl; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00179; EGF_Ca; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE: PS0240; TRYPsin_DOM; 1.
DR PROSITE: PS00135; TRYPsin_SER; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydrolase; Lectin;
KW Protease; Repeat; Serine protease.
SQ SEQUENCE 719 AA; 80425 MW; 17064151870444E5 CRC64;

```

Query Match 3.1%; Score 137.5; DB 13; Length 719;
 Best Local Similarity 19.5%; Pred. No. 0.0045;
 Matches 100; Conservative 62; Mismatches 149; Indels 201; Gaps 27;

```

OY 235 IMSANYPEGPEDELMTQGVVPAHRLASVFLNPNLSN-----CE-----RKEEREYX 284
Db 29 IQSPNTPPEGIPSLDVTWNTTTPAGRIKILYFHHFDLEPYLCEIYVYKVSDELLER 88
OY 285 IPGSTNPEYFKLEDKOPGNMAGNPNLSLOGCDQDQASPGILRLQVLYVHPDONSNKI 344
Db 89 CGRESTDTE-----QIRP-----RE 103
OY 345 YVVDLSNERAMSLTTPRPYKOSRKFVPGCFVLESRTCSNLTLSGSKHKSIFLCLDL 404
Db 104 EV--LINSNTMAVTLTKTDFSNBER-----FTGEFAHYSAVDID-BCEVKIDALVCDHE 154
OY 405 TRLMNVETISC-----TDHRYCO-----RKSYSLOVP-----SDILHL 439
Db 155 CHNYIG-GYICSCRFSTYILHTDNRTCKVECDNVFTKSGVISPDPNPPSSDCLY- 212
OY 440 PVELHD-FSWKLLVPKRLSLVLPAAKLOOHTHERKPCNTSFSYLVASAIPSDLYGSP 498
Db 213 RILEEGFA-----ISLEFADMEFLEDDHP-DVPC--PYDYL--KIKAGSMAGPFP 257
OY 499 CPBGSLKQIOVKONISVTLTFTAPSFQOENSRQGLVSTF-----PYKEEGVFTVP- 551
Db 258 CGDKSPGKIDTNS-SVQILFHSG--SGDNQGMKMSFSLVGTGPIILHPANGTIEPV 313
OY 552 ---DTRSKYLTLPMMNDRLPSLTSV-----SWNISVPRDOVA-C-----587
Db 314 QDOYSKRDVQIKCNIGFKFKVAGDEYAVQICQXSGSSSIPCKYIDCCAPPLED 373
OY 588 -----LTFEKRSGVVCOTGRAFMIIQORTAEETFSLDEVDLPKPSFHHR--- 634
Db 374 GITFHSKENVLFRSHIEVSC-----PALHHLER 404
OY 635 -----SWVN-----ISNCSPTSGK 649
Db 405 ELNWTYTCMENGFWVDELGTDLPTCOPVCGR 436

```

RESULT 9

```

O70244 PRELIMINARY; PRT; 3623 AA.
ID O70244;
AC O70244;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Intrinsic factor-B12 receptor precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=98148073; PubMed=9478979;
RA Meistrup S.K., Koziyaki R., Kristiansen M., Kayen J.H.,
RA Rasmussen H.H., Brault D., Pontillon F., Gode F.O., Christensen E.I.,
RA Hammond T.G., Verroust P.J.;
RT "The intrinsic factor-vitamin B12 receptor and target of teratogenic
RT antibodies is a megalin-binding peripheral membrane protein with
RT homology to developmental proteins.";
RL J. Biol. Chem. 273:5235-5242(1998).
CC -I- SIMILARITY: CONTAINS 26 CUB DOMAINS.
DR EMBL: AF022247; AAC71661.1; -.
DR HSSP: P00740; IEDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR00742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00431; CUB; 27.
DR Pfam: PF00008; EGF; 7.
DR SMART: SM00042; CUB; 26.
DR SMART: SM00179; EGF_CA; 4.
DR SMART: SM00001; EGF_like; 4.

```


DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE Silec-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21523976; PubMed=11546777;
 RA Angula T., Varki N.M., Varki A.;
 RT "A second uniquely human mutation affecting sialic acid biology";
 RL J. Biol. Chem. 276:40282-40287(2001).
 DR EMBL: AF282256; AAK71521.1;
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig_4.
 DR SMART: SM00406; IGV. 2.
 SQ SEQUENCE 595 AA; 64984 MW; D05662176274C5C3 CRC64;

Query Match 2.8%; Score 126; DB 4; Length 595;
 Best Local Similarity 19.9%; Pred. No. 0.035;
 Matches 135; Conservative 86; Mismatches 247; Indels 212; Gaps 34;

OY 225 SVEEGESATLMSANYPGEFPEDELTMTQFVPAH---LRASVFLNLSNCRKEERY 281
 DB 33 TQEGCLCVSLCSFSPON-----GWTASDPVHGWTWRAG-----DHV 70
 OY 282 EYTPGSTNPEVFKLEDKOPGNMAGNFNLSLOGCDODASPGILRQFVLQHPONES 341
 DB 71 SRNIPVATNNPARAVOEE-----TDRFR-LIGDQONKO 103
 OY 342 NKIIYVDLSNERAMSLTIERPRYKOSKRVPGCFV-CLSRCTSSNLTLSGKHISFL 400
 DB 104 CTL-----SIRDRESADAGTYVFCVERGNMKMYKKYDQLSVN-VTAS 144
 OY 401 CDDLTRLMNVKERTISCTDHRVCORSYSLOVPSDLH-----LPV-----ELHDE 446
 DB 145 QDLSTKRLLEVPSYV-----QEGLCVSPCSVLPIRYMKTASSPYGSKFKGADI 197
 OY 447 SKMLLVPKDRLSLVPAQKLODHTH-----EKPCNTSFSYVASAIPSDLYGSGF 498
 DB 198 PMDIPVATN-----TPSGKVQEDHGRFLLDGDPQTNNCSLSIRARAGDSGKYYFO-- 249
 OY 499 CEGGSK--QIOVKONISVTLRTFAPSFO-----QASROGLVSTPIYFKEEG----- 545
 DB 250 VERGSRKNMYIYDKLSVHTALTALHMTPTFSIPGLESCHPRNLCS-VPMACEQGTPTIT 308
 OY 546 -----VFTVTPD-TKSKVYLTRPN-WDRGLPSLT-----SVSNMISVPRQ 584
 DB 309 WNGASVSLDPTITSSMSLSLIPQODHG-TSLTCOVTLPGAGVTMTRAVRLNISYPPON 367
 OY 585 VACLTFEKE-RSGVVCOTGRAFMIIQORTRAEIFSLDE-----624
 DB 368 LMTVVGQGGTASTLRNSALSYLEGQSLHL--VCAVDSNPRARLSMTWMTGSLTSPSS 425
 OY 625 -----VLPKSFH---HNSWVNISNCSPTSGKOLDLFSV--TLPRVDLVILAAVG 675
 DB 426 SNLGELEPRVHKDEGEFTCRAN--PLGSHISLSLSKONEYTKMMPISGVTLGARG 483
 OY 676 GG---VLLSALGLICCVKKKKKKTKNKGPAVGYNGNINTERPRPKFKOKRKNDSH 732
 DB 484 GAGATALLVLYFCILFVVVVRSCRRKSAR-PAVGV--GDTGME-----522
 OY 733 VYAVIEDTVYGHLLQDSSGSPLOPEVDVYREFQGTMGVCPSPPTICSHAPAKLATER 792
 DB 523 -----DANAVRG---SASGRLIESPADDSRPHARPALATSPREE--GEIQAASLSFK 572
 OY 793 PPRPRPSESEPTYTFSHPN 812
 DB 573 ARQYPOEOEAIGYEYSEIN 592

RESULT 12

076355

PRELIMINARY; PRT; 986 AA.

AC 076355;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE C45G7.5 protein.

GN C45G7.5.

OC Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Bonfield J., Alnsough R., Anderson K., Baynes C., Berks M.,

RA Craxton M., Dear S., Du Z., Durbin R., Favallo A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,

RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Dante M., Wamsley P.;

RT "The sequence of C. elegans cosmid C45G7.";

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RP [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF067611; AAC19183.1;

DR InterPro: IPR002126; Cadherin.

DR PROSITE: PS50268; CADHERIN_2;

SQ SEQUENCE 986 AA; 109455 MW; 80729CDE64D9B93 CRC64;

Query Match 2.8%; Score 123.5; DB 5; Length 986;
 Best Local Similarity 18.3%; Pred. No. 0.13;

Matches 128; Conservative 102; Mismatches 242; Indels 229; Gaps 32;

OY 127 WDVKAKSIGLELOFSIPR-----LRQIGGEGCPDG 158
 DB 387 FETKNNMAVACELHWSVPRKRSQLECRLEKILQALRENNKRLYLRLGMHGKT-PVK 445
 OY 159 YTHSISGRIDATVVRIGTSPCSNCTVSRIKQGVKALHLPWHPRVVSFSTANSSIK 218
 DB 446 LAES-----STVENISKDREDSVKGITYRSYTR--LPYLKKTPEWS----- 488
 OY 219 RLCTISVFEGESATLMSANY-----PEGFPD-----ELMTQFVPAHLRASVPLNF 269
 DB 489 QYITLQD---EKSDVESNYPIDQRPQKFLDTEPAELLITAFNSGNLTVSSITLTKI 545
 OY 324 GI-LRQFOVLQHPQ-NBSNKIYVVDLSNERAMSLTIERPRYKOSKRVPGCFVCLSR 381
 DB 606 NFNIIDAEFEIYEPKVDPSKMTAEV-----CFIATSOQ 641
 OY 382 TC-SNLTLLTSGSKH---KISFLCDLTRLAMNVKERTISCTDHRVCORSYSLOVPSDI 436
 DB 642 NLDTSFVFPDANSKNKRVKVEFGD-----SVDSVANSGRINNGFSGNQSISSVGDV 694

```
QY      437 L-HLPEL-----HFSKLLVPRDRSLVLPVPO-----KLIOO THE 473
D        | : : | | | | | | | | | | | | | | | | | | | |
Db      695 LDREIPIPNILAAAGADYEOK-----SLNVVPANTELGRDMISPEGAVELNKKTG 746
QY      474 KPCNTSESYLVASA---IPSODLY---FGSHCPGSGIKOIUVON-----ISVLT RTE
D        ||| : | : | : | : | : | : | : | : | : | : |
Db      747 -----LVANKRIIDTPOGYTTAELGRDGEGYKQLHIIRNDKRLXYVLSMSNEF 757
QY      521 APSFOQEASRGLTVSFIPYFKKEGVFTVPDTKS-----KYVLRTPNMDRGLEPSLTSVS
D        | : : | | : | : | : | : | : | : | : | : | : |
Db      798 GANLEK-----FRQILIEAGMDQKAKQLEIHDEFEKADRKNSTWTSV- 842
QY      576 WNIYSVRDVOACLTFFEKERSGVQCOTGRAFIITOEORTRAEETSLDEDVLPKRPFHHHS 635
D        | : | : | : | : | : | : | : | : | : | : | : |
Db      843 -----CFYLTRQNAIIDENQASSIISSNGHISKL-----HHI 875
QY      636 FMV-NISNCSP-----TSGRQDLDSVTLTPTVDLTYLIAAGVGULLSALG 685
D        | | : | : | : | | | | | | : | : | : | : | : |
Db      876 FKVOAWDMCAKPBATSESTSSNSDPLN-----TLTLIGVGLLIALLAL 925
QY      686 LIICCVK-----KKKTNK---GPAVGIVGNINTEMPROP 719
D        | : | : | : | : | : | : | : | : | : | : | : |
Db      926 IYVCVSRYQRYLKQKTDQLRCSSSASASYKSPMLIPPPPP 966
```

RESULT	13
09T053	
ID	09T053
AC	09T053; PRELIMINARY; PRT: 3620 AA.
DT	01-MAY-2000 (TREMBlrel. 13, Created)
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	Cub11n.
GN	CUBN.
OS	Canis familiaris (Dog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX	NCBI_TaxId=9615;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=KIDNEY;
RX	MEDLINE=20021710; PubMed=10552972;
RA	Xu D., Kozyraki R., Newman T.C., Fyfe J.C.;
RT	"Genetic evidence of an accessory activity required specifically for
RT	cub11n brush-border expression and intrinsic factor-cobalamin
RT	absorption";
RL	Blood 94:3604-3606(1999).
CC	-1- SIMILARITY: CONTAINS 26 CUB DOMAINS.
DR	EMBL; AF137068; AAF14258.1; -.
DR	HSSP; P35555; 1EKN.
DR	InterPro: IPR000152; Asx_Hydroxyl.
DR	InterPro: IPR000859; CUB_domain.
DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR000742; EGF-2.
DR	InterPro: IPR001881; EGF_Ca.
DR	Pfam: PF00431; CUB; 27.
DR	Pfam: PF00008; EGF; 6.
DR	SMART; SM00042; CUB; 26.
DR	SMART; SM00179; EGF_CA; 4.
DR	SMART; SM00001; EGF_like; 4.
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.
DR	PROSITE; PS01180; CUB; 27.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR	PROSITE; PS01186; EGF_2; UNKNOWN_2.
DR	PROSITE; PS01187; EGF_CA; 4.
RW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ	SEQUENCE 3620 AA; 397429 MW; 00B041EE6AD07348 CRC64;
Query Match	2.8%; Score 123.5; DB 6; Length 3620;
Best Local Similarity	18.1%; Pred. NO. 0.94;
Matches 154; Conservative 104; Mismatches 285; Indels 309; Gaps 40;	

QY	51	LLAPRCYIVISKRIITWLSIKSEGRIVTFECSQSPENHVEILOKNI--DC--MSGPCF	106
Dd	2289	LLALIC-----GKSPBSQJSSSEWYVWLRFRSDNSSTQVGFYKIKALIAOCCGRVGTGSGI	2343
QY	107	GEVLOPSTSLPTLNTFT--WDVKAHKSIGLELOFSIRLROIP-----	151
Dd	2344	IESGCEPT---LPRDNSFCFEMHLKPCSGHYILTFHEDEHLONSSGCEKDFEIVEMENTS	2400
QY	152	-----GEGCPGVHHSISGRIDATVVALGTFECSNGYSTRKMOEGVKMALHLPWFHPRVVS	207
Dd	2401	GMLGRYCGNTIPDSITSSNALVR---FVYDGSYT-----AS	2436
QY	208	GFSIANRNSIKRLOLIESVEGE---GSATLMSANPEGFPDELMTQFVYPAHLRASV	264
Dd	2437	GFRLRFESSME-----ACGGELQOPRTGTFSPNPNPNHGRVCEMRIMWOGBSRITLL	2489
QY	265	SFLNPNUS---NCERKEEVEYIIPGSTINPEVEFKEDKOPONMAGNFNLSIQGCCDOAQ	321
Dd	2490	TFNMLRLERAHPSG--YSEHTITNGIRNNSPOLEKL-----CGSVAS---SEIK	2534
QY	332	SPGILRLQFVLVQHPONEBKNLYVDLSNERAMSLTEPRVQOSRKF--VPGCFVYLE	379
Dd	2535	SSG-----NIMKVVFFPD-----GSRPGGGSATYTSSE	2563
QY	380	SRTGSSNLT-----LTSGSKHKISFLCDLRLMNNVEKITS-----	416
Dd	2564	DAYCGGSLTRHPREGNFTSPGYNVSYSRNL-----NCEMTLSNPNOCNSIYIHFEFVY	2618
QY	417	CTDHRVQOKRSYSLQVPSDLHLRVLEHD--PSMKLYPKRSLVLVPAOKL-----Q	468
Dd	2619	LESHQDCQ-----FDVLEFRVGNAGDPLMMKRLCGFSKFIYVPLVLYPEVMWHEVYN	2669
QY	469	QHTHEKPCNTSFSY-----LVASAPISSODLYFGSFCGGSIKO---IOYKONISVTL	517
Dd	2670	EHEVHGFAHEVSTDCGGIGLOGESVYIASPNYPASV---DSLTHCSMLLEAPQGFITLL	2726
QY	518	RTFA-----PFOEDASRQGLTYSFIYF	541
Dd	2727	-TFDEFDIEDHATCAMDSVSVNRNGSGPSPDIIGQYCGTSPNPTTIOGSGMQLVVFNSDHS	2785
QY	542	KEEVEVTFVTPDTS-----KUYLRTPMNDRGLPGLTGSVW-----	576
Dd	2786	VONGGFATNTNQILOGCGGLIHSNDNTIISPNNHPQVFNPSNCSMTVYITHESKOLEISFD	2845
QY	577	-NISVPRDOYACLTFP-----KERSGVVOCOTG-----RAFM--TIOEQTR	614
Dd	2846	NNFRIPSGDGCONCSFYKAGTAEVEAESILATGCGNAPAGSILTRPRVFIWFOGSETP	2905
QY	615	AEEIFESIDEDVLRKPSFHHHSFVWNJSNC-----SPTSGKOLDLFSVTLPR	662
Dd	2906	AQG-FSA-----SF---VSRGCGNFTNPSGILTSPNTPROYDNMNTIYIE	2948
QY	663	TVDLTVLILI-----AAVGG-----GVLLSALGI-----ITCCVKKKKKKTKNG	701
Dd	2949	ADPLSVVILTFEESFHEARSAITGSCANDGVHILIRGNSLNSPRTAVCGNETLSPVTLIG	3008
QY	702	PAGVITYNGINTF 713	
Dd	3009	PVLLNFTSNAHT 3020	

RESULT	ID	PRELIMINARY:	PRT:	724 AA.
093VB7	093VB7			
AC	093VB7;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	P0487H02.1 protein (OSUBA0025P13.19 protein).			
GN	P0487H02.1.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
CC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			

OC Euhartolidae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC clone: P0487H02."
 RN Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, BAC clone: OSJBA0025P13."
 RN Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002883; BAB6784.1;
 DR EMBL: AP003140; BAB5759.1;
 DR InterPro: IPR001525; C5_DNA_meth.
 DR PROSITE: PS00094; C5_MTASE.1; UNKNOWN.1.
 SQ SEQUENCE 724 AA: 79780 MW: D6B3E09BEAF22253 CRC64;

Query Match 2.8%; Score 123; DB 10; Length 724;
 Best Local Similarity 21.3%; Pred. No. 0.088;
 Matches 114; Conservative 61; Mismatches 201; Indels 160; Gaps 27;

OY 348 DLSNRAMSLTEPRPV--KOSRKYPGCGVCLSESTCSNLTLSGSKHKISFLCDDLT 405
 DB 260 ELQRDACLGLGHRITADQERERP-----RSEESTTPRSISLPLPSMADQLR 310
 OY 406 LWM-----MNVEKTSIC-TDHRYSQKRSYSLQVPSDILHLPVLDHFSWKLVPKD- 455
 DB 311 RMAAADOGLLEVARICELDRHMGVCSMRALYKLAAPRPPPLPALPESD 370
 OY 456 ----RLSLVLPAAKLOQHTHEKPCNT-SFSTYLVASAIPSODLYFGSGFCG-----GSI 504
 DB 371 GLPAPVSCVL-----SGCRTHAFSVLQGA--RGARYFGSYDGMFLAVGGO 415
 OY 505 KOIOYKONISVT-----LRTFAPSFQOEGSRGLVSPFPRKEE-----GVF 547
 DB 416 AOCQALLNKLINGFOTLPLNLAIPVSNPNDRREMAIVAAULTSCPP--TECGCIYAGII 473
 OY 548 TVTPD-----TKSKVYLRTPNMDRGLPSLTSSVSNISVPRDOVACLTFFKERSGVQO 600
 DB 474 ESSPMLVANGHTSLAF-----WRMGDQVLPVLMAL-----EEDNPLMRLEVEDLGH 524
 OY 601 TGRAMIIIOBORTRAEELFSLDEDLVLPKPSFHHSHFWNISNCSPTSGKQDLLESVTLT 660
 DB 525 HG-AEHFL--TRAEDVLACEE--PRVFYRDSVSLVPAN-----MFL 561
 OY 661 PRTVP--LTVILIAVGGVLLLSALGLTICCVKKKKKTKNGRAVGITNGINTEMPRQ 718
 DB 562 PNVHDENEVLAHYLVGSKLL--MVVRLASGRGRTTS--AFRYV-----Q 605
 OY 719 PKFKQKGRKNDSHVAVIEDTMYGHLLODSSGFLQAEVTVYPRFOGTWGVCP----- 773
 DB 606 KKKFTGEEDDESQNSRAFE--YTWSELDLDGRMLTFLGLCSYKAGDGRYEMERG 663
 OY 774 -----PSPPTIC-----SAPTAKLATEE-----PPRSP 799
 DB 664 VFLLDPSIHQWIGDAPKPYLCSDNCKMSRAPTPDQGOVRCRPERPSIHSP 719

RESULT 15
 OY9216 PRELIMINARY; PRT; 1406 AA.
 AC OY9216; O9NOE3;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DE KIA00980 protein (D0691N24.1.1) (KIA00980 protein, isoform 1) (fragment).
 GN KIA00980 OR D0691N24.1.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=BRAIN;
 RX MEDLINE=99245063; PubMed=10231032;
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."
 RL DNA Res. 6:63-70(1999).
 [2]
 RP SEQUENCE OF 22-1406 FROM N.A.
 RA Blakey S.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB023197; BAA76824.1;
 DR EMBL: AL031672; CAB99291.1;
 DR HSP: P02633; 4ICB.
 DR InterPro: IPR02048; EF-hand.
 DR Pfam: PF00036; efhand; 3.
 DR SMART: SM0054; Eph; 2.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN.1.
 FT NON-TER 1
 SQ SEQUENCE 1406 AA: 158951 MW: A530020275943655 CRC64;

Query Match 2.7%; Score 118.5; DB 4; Length 1406;
 Best Local Similarity 17.8%; Pred. No. 0.61;
 Matches 154; Conservative 120; Mismatches 306; Indels 283; Gaps 37;

OY 133 KSLGL-----ELQSTIPRLQIGPESCDGVTSHISGRIDATVYRIGTSGNGTVSRI 186
 DB 250 QSVGLGLKEKELEDELNFNLQDDG-----DCKVSLSEFQLGIFSHR--PAL 293
 OY 187 KMOEGVKMALHLHPFHR--NVSGFSIANRSIKRLC-----IIESVEEGSGATLSANY 240
 DB 294 LLESTRVPRSKRMSHYQVPEESGCHTTTSSLVSLCSSLRFLSSIDDSGRA----- 346
 OY 241 PEGFPEDELTWQFVVAHLRASVSFLNLSNCKEREVEYYIPGSTNPVEVFLDK 300
 DB 347 ----FPQVLAAMTQEGIQNGREILQSLDSV-----DEKVNLL----- 381
 OY 301 QPQNNAGNPLSLQCGDDQAQSPGILRLQFVLVQHPQNSKITYVVDLSNERAMSLTE 360
 DB 382 ---ELTWALDNEIMLTVDASVVOQAALACYHQEELSYQGG-----VEQLAREDKARQDL 431
 OY 361 PRPVKSRKFV--PCGFVCLSESRGSSNLTLSGSKHKISFLCDDLTFLMMNVEKTSIC 417
 DB 432 ERAEKNLEFVEMDDCHSTLEQLTEKKIKHLEQGRERLSTL--RSEVAERELFW 486
 OY 418 TD-HRYQKRSYSLQVPSDILHLPVLDHFSWKLVPKDRSLVLPAAKLOQHTHEKPC 476
 DB 487 EQAHR--QAAA-----LEMYVGRLOAEAGL-----RKLTALKENSRLOKEIYEVE 533
 OY 477 NTSFSTYLVASAIPSODLYT--GSFCGGSIKQIQVKNISVTLRTFAPFQOBASQGLT 534
 DB 534 KLDSERLALKL-QKDLFVLKDKLEPQSAEILQAEERAAVLYKEYELKCRDLQDRNDEL 592
 OY 535 VSPITPFKEGEGVTVYPRKSKVYLTPNWD-----RGLPSP----- 571
 DB 593 QAEL-----EGIMARLPKRR-----HSPWSPDGRRRQLPGLAPAGISFLGNSAPYSIT 642
 OY 572 -----TSVSMNISVPRDOVACL--TFKERSGVVCTGTFAMIIQDORT 614
 DB 643 ELAMBEQVKEHYQDLRTQLFTKYNYERETALAKRNEKERKMQEQRREYVLSQKAD 702
 OY 615 AEETISLDDVL-----PKP-----SFHHSTFWNINSCSP 645
 DB 703 LELHEKSOEVIWGOLODQDTARGEPEPMGLAPCCTALGGLAHRHSHSHQOIRREAE 762
 OY 646 T--SGKQDLRLSVTLTPRTVDLTVILIAVGG-----GVLLSALGLTICVVK 693

```
Db 763 AELSGE---LSGLGALPARDLTLELEPPQGPLPRGSGRSEQLELERALKLQPCASEK 818
QY 694 KKKTKNGPAVGIIYNGNINTEMPQPKKFOGGRKN---DSHVYAVIEDTWYGHLLQD 749
Db 819 RAQMC-----VSIALFEEELRLAKRKRYDGPISLEAMQALPKDGLVAG----- 861
QY 750 SSGSEFLQPEVDIYRPFQGTWVCPPSPPTICSRAPTAKL----- 788
Db 862 -SGQ-----EGTRGLPLRPG--CGERPLAWLAPGDGSESEEAAGAPRRQA 906
QY 789 ---ATEEPPRRSPPESE--SEPYTFSH--NNGDVSSKD-----TDIP 824
Db 907 QDTEATQSPAPAPAPASHGSPSRMSRMQPCGVGDIVPKEPEPPGASAGLEQPGARELP 966
QY 825 LLNTOE-----PMEPA 835
Db 967 LLGTERDASQTOPRMWEPPLRPA 989
```

Search completed: November 11, 2002, 11:35:52
Job time : 88 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2002, 03:25:40 : Search time 15168 Seconds

(without alignments)
11824.934 Million cell updates/sec

Title: US-09-899-569a-3

Perfect score: 6163

Sequence: 1 ccaagccgcgcattggggagtc.....agtcgcgcaaaaaaaaaa 6163

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6163	100.0	6163	6 AX353605	AX353605 Sequence
2	5963	96.8	5963	9 AY026461	AY026461 Homo sapi
3	5854.2	95.0	5897	6 AX353603	AX353603 Sequence
4	5518.4	89.5	5573	9 AK023834	AK023834 Homo sapi
5	3769.8	61.2	163115	9 AC105902	AC105902 Homo sapi
6	3769.8	61.2	200956	9 AC104165	AC104165 Homo sapi
7	3201.6	51.9	3224	9 AK026622	AK026622 Homo sapi
8	1875.8	30.4	2113	9 AK026028	AK026028 Homo sapi
9	1232.6	20.0	1241	9 AK026187	AK026187 Homo sapi
10	1159.8	18.8	1410	9 BC021099	BC021099 Homo sapi
11	1117.8	18.1	1376	6 AX464028	AX464028 Sequence
12	959	15.6	976	9 AK026329	AK026329 Homo sapi
13	821.2	13.3	3843	10 BC011340	BC011340 Mus muscu
14	619.4	10.1	1999	10 BC034137	BC034137 Mus muscu
15	542.4	8.6	544	6 AX192795	AX192795 Sequence
16	533	8.5	541	6 AX261684	AX261684 Sequence
17	522.6	8.5	636	6 AX261015	AX261015 Sequence
18	482	7.8	535	6 AX260742	AX260742 Sequence
19	380.4	6.2	400	6 AX203285	AX203285 Sequence
20	364	5.9	173977	9 AC010170	AC010170 Homo sapi
21	304.8	4.9	157126	2 AC121182	AC121182 Rattus no
22	304.8	4.9	179992	2 AC095165	AC095165 Rattus no
23	247.4	4.0	229379	9 AC010746	AC010746 Homo sapi
24	245	4.0	195768	9 AC022034	AC022034 Homo sapi
25	244.4	4.0	125517	9 HS23061	HS23061 Human DNA
26	242.6	3.9	191035	9 AC093527	AC093527 Homo sapi
27	239.6	3.9	230516	9 HS052111	HS052111 Homo sapi
28	239	3.9	172577	2 AL365174	AL365174 Homo sapi
29	238.6	3.9	40639	9 AC026805	AC026805 Homo sapi
30	238.6	3.9	44724	9 AC124224	AC124224 Homo sapi
31	238.4	3.9	171673	2 AC073470	AC073470 Homo sapi
32	237.8	3.9	172515	9 AC096565	AC096565 Homo sapi
33	237.6	3.9	178670	2 AC009290	AC009290 Homo sapi
34	237.6	3.9	163026	2 AC103883	AC103883 Homo sapi
35	237.6	3.9	175054	9 AC010531	AC010531 Homo sapi
36	237.4	3.9	78947	9 AL136126	AL136126 Human DNA
37	237.4	3.9	157784	9 AL360232	AL360232 Human DNA
38	237.2	3.8	124388	9 AL161457	AL161457 Human DNA
39	237.2	3.8	152961	2 AC025021	AC025021 Homo sapi
40	237	3.8	159699	2 AC026833	AC026833 Homo sapi
41	237	3.8	188295	9 AC098869	AC098869 Homo sapi
42	236.8	3.8	171550	2 AC127004	AC127004 Homo sapi
43	236.8	3.8	177829	9 AC010178	AC010178 Homo sapi
44	236.4	3.8	89000	9 AP003059	AP003059 Homo sapi
45	236.4	3.8	101766	2 AP000632	AP000632 Homo sapi

ALIGNMENTS

RESULT 1
AX353605
LOCUS
DEFINITION Sequence 3 from Patent WO0204508.
ACCESSION AX353605
VERSION AX353605.1 GI:18618678
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
1 Schweifer,N., Scherl-Mostagier,W., Sommergruber,W. and Abseher,R.
Tumour-associated antigen (b45), characterised by an amino acid
sequence as in seq. Id. No. 4

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0204508-A 3 17-JAN-2002;
FEATURES Location/Qualifiers
source 1..6163
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1..282
GC_signal 147..157
misc_feature 201..209
/note="cap signal: Transkriptionsstart"
283..2793
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD22954.1"
/db_xref="GI:18618679"
translation="MAGLNCGVSIALLGVLLGARLRPGAEAFELALPRESNTVLI
KLGPTPLAKPCYIVISRHITMISIKSIVTEFCOSPENHEVIEKNIDMSG
PCPEGEVLOLPSTSLPTLNRTFIMDKAKISIGLEOFISIPRLROIGPESPDPYGT
HISGRIDATVVRITGTCNSGTYSRIKMQGKMAHLTFPHFRNNSGESIANRSIK
RLCTIESVEGESATLMSANYPGEFEDLMTQEVPAHLRASVFLFNLSNCR
QHPONESNKIYVDLSNERAMSLTIEPRPKOSKFPVGCFCLESSTCSNLTLSG
SKHISFLCDLRLMMNVKTSCTDHRKORSYLOVPSDILHLPELHDSMKL
LVPRDLSTLVPAOKLOOHTHEKPCMTSFSYLSAIPSDILYFSGFCGSGSIKOJO
VKNISYTLTEFAPASROEASROGRTYSFTPYKEBVFVTPDKSKYLYLRPNMR
GLPSLVSNNISVPRDVACLTFFKRSVVCOTGEAFNIIDORTAREITSLEDD
VLPRSPHSHSEFWNISNCSPTSGKQDLLESVTLTPRTVDLVILIAAGGVLLS
ALGLIICCVKRRKTKNKPAVGIVNGNITEMPRPKROKGRKNDSHVAVIEDT
MYEGHLDDSSGSFLQPEVDYRPFQGTMCVCPSPPTICSRAPTAKLATEEPPRSP
PESESEYTFSSHNNNGDVSFKDTDIPLNTQEPMEPAE"
3'UTR 2794..6163
BASE COUNT 1501 a 1689 c 1482 g 1491 t
ORIGIN
Query Match 100.0%; Score 6163; DB 6; Length 6163;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCAACGCCGCAATGGAGTAGTAGGGAACCCAGCAACCCGGTGCCGGAGCCCTGCACCC 60
DB 1 CCAACGCCGCAATGGAGTAGTAGGGAACCCAGCAACCCGGTGCCGGAGCCCTGCACCC 60
QY 61 TGGAGAGGAGAGGGGGTGGCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 61 TGGAGAGGAGAGGGGGTGGCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 CGGGTCAGCTGCGATCTGCTGCGCAAGGGGCTGCGGCGGTCGCCGCCGCGCGAG 180
DB 121 CGGGTCAGCTGCGATCTGCTGCGCAAGGGGCTGCGGCGGTCGCCGCCGCGCGAG 180
QY 181 GTGAGTAGACCGAGGCGGAGCGAGCTGCGCGGCTTGGGCGCGCTGGGCGCGCTCC 240
DB 181 GTGAGTAGACCGAGGCGGAGCGAGCTGCGCGGCTTGGGCGCGCTGGGCGCGCTCC 240
QY 241 CCACCGTGTCTTCCCAACGAGGCGGAGGCGTCCGGAGTCATGGCGGCGCTGAACGCG 300
DB 241 CCACCGTGTCTTCCCAACGAGGCGGAGGCGTCCGGAGTCATGGCGGCGCTGAACGCG 300
QY 301 GGGGTCTCTATCGCACTGCTAGGGGTTCTGCTGCTGGGTGCGGCGGCTGCGCGGGG 360
DB 301 GGGGTCTCTATCGCACTGCTAGGGGTTCTGCTGCTGGGTGCGGCGGCTGCGCGGGG 360
QY 361 GCAGAGAGCTTTGAGATTGCTCTCCACGAGAAAGCAACATTACAGTCTCTATAAGCTG 420
DB 361 GCAGAGAGCTTTGAGATTGCTCTCCACGAGAAAGCAACATTACAGTCTCTATAAGCTG 420
QY 421 GGGAGCCCGAGCTGCTGGGAGAAACCGTGTACATCGTCATTGTTAAAGACATATAACC 480
DB 421 GGGAGCCCGAGCTGCTGGGAGAAACCGTGTACATCGTCATTGTTAAAGACATATAACC 480
QY 481 ATGTTGTCATCAAGTGTGAGAAAGATAGCTTTTACCTTTAGCTGCACAGTCTGAG 540
DB 481 ATGTTGTCATCAAGTGTGAGAAAGATAGCTTTTACCTTTAGCTGCACAGTCTGAG 540

QY 541 AATCACTTGTTCATAGAGATCCAGAAAAATATGACTGTATGTACAGCCCATGTCCTTTT 600
DB 541 AATCACTTGTTCATAGAGATCCAGAAAAATATGACTGTATGTACAGCCCATGTCCTTTT 600
QY 601 GGGAGAGTTCAGCTTCCACCCCTGCACATCGTGTGGTCCCTACCCTCAAGAACTTCAATC 660
DB 601 GGGAGAGTTCAGCTTCCACCCCTGCACATCGTGTGGTCCCTACCCTCAAGAACTTCAATC 660
QY 661 TGGAGTGCAGAAAGCTCATMAAGAGATCGGTTAGAGCTGACAGTTTCCATCCCTCGCTG 720
DB 661 TGGAGTGCAGAAAGCTCATMAAGAGATCGGTTAGAGCTGACAGTTTCCATCCCTCGCTG 720
QY 721 AGCGAGATCGTCCGGGTGAGAGCTGCCAGACGAGTCACTACTCATACAGGCGCGA 780
DB 721 AGCGAGATCGTCCGGGTGAGAGCTGCCAGACGAGTCACTACTCATACAGGCGCGA 780
QY 781 ATCATGTCACCGGTGTGAGATGGAACCTTTCACACATGGAACGTGACGTGTCCCGGATC 840
DB 781 ATCATGTCACCGGTGTGAGATGGAACCTTTCACACATGGAACGTGACGTGTCCCGGATC 840
QY 841 AAGATGCAGAGAGAGTGAATAATGCGCTTACACCTCCCATGTTCCACCCAGAAATGTC 900
DB 841 AAGATGCAGAGAGAGTGAATAATGCGCTTACACCTCCCATGTTCCACCCAGAAATGTC 900
QY 901 TCCGCTTCAGCATTCGCAACCGCTCATCTATAAAGCTGTGTGCATCATGAGTCTGTG 960
DB 901 TCCGCTTCAGCATTCGCAACCGCTCATCTATAAAGCTGTGTGCATCATGAGTCTGTG 960
QY 961 TTTGAGGGTGAAGGCTGCAACCGCTGATGTGTGCAACCTCCAGAGAGGCTTCCCTGAG 1020
DB 961 TTTGAGGGTGAAGGCTGCAACCGCTGATGTGTGCAACCTCCAGAGAGGCTTCCCTGAG 1020
QY 1021 GATGAGCTCATAGAGTGGCACTTTGTCTCTGTCACACCTGCGGGCGAGGCTCTCTTC 1080
DB 1021 GATGAGCTCATAGAGTGGCACTTTGTCTCTGTCACACCTGCGGGCGAGGCTCTCTTC 1080
QY 1081 CTCAACTTCAACCTCTCTCAACTGTGAGAGAGAGAGAGCGGTTGAATTAATCAATCCCG 1140
DB 1081 CTCAACTTCAACCTCTCTCAACTGTGAGAGAGAGAGAGCGGTTGAATTAATCAATCCCG 1140
QY 1141 GGCTTCACACCAACCCCGAGGAGTGTCAAGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 GGCTTCACACCAACCCCGAGGAGTGTCAAGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GGGAACTTCAACCTCTCTCTCAAGGCTGTGACCAAGATGCCAAAGTCCAGGATCTCTC 1260
DB 1201 GGGAACTTCAACCTCTCTCTCAAGGCTGTGACCAAGATGCCAAAGTCCAGGATCTCTC 1260
QY 1261 CGGCTGCAGTTCCAAGTTTGTGTCACACATCCACAAAATGAAGCAATMAAATCTACGTG 1320
DB 1261 CGGCTGCAGTTCCAAGTTTGTGTCACACATCCACAAAATGAAGCAATMAAATCTACGTG 1320
QY 1321 GTTGACTGTGAATGAGAGAGGATGTCATCAACATCGAGAGCCAGGCGCTCAAAAG 1380
DB 1321 GTTGACTGTGAATGAGAGAGGATGTCATCAACATCGAGAGCCAGGCGCTCAAAAG 1380
QY 1381 AGCGCAAGTTTGTCCCTGCTGCTGTGTGTAGATCTCGAGACCTGAGTAGCAAC 1440
DB 1381 AGCGCAAGTTTGTCCCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
QY 1441 CTCACCTGACATCTGGTCCAAACACAAAATCTCTCTTGTGTGTGTGTGTGTGTGTGTGT 1500
DB 1441 CTCACCTGACATCTGGTCCAAACACAAAATCTCTCTTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 1501 CTGTGATGATGTGAAAAAACCATTAAGTGCACAGACGACGAGTACTGCAAAAGAA 1560
DB 1501 CTGTGATGATGTGAAAAAACCATTAAGTGCACAGACGACGAGTACTGCAAAAGAA 1560
QY 1561 TCTTACTACTCCAGGTGCCCCAGTGAATCTCCACCTGCTGTGTGTGTGTGTGTGTGTGT 1620
DB 1561 TCTTACTACTCCAGGTGCCCCAGTGAATCTCCACCTGCTGTGTGTGTGTGTGTGTGTGT 1620
QY 1621 TCTGTGAAGCTGTGTGTGCCAAGGACAGGCTCAAGCTGTGTGTGTGTGTGTGTGTGT 1680

D	b	1621	TCTGTGAAAGCTGCTGTGTGGTCCCAAGSACAGGCTCAGCGCTGTGTGTGTGCCAAGCCAGNAG	1680
O	y	1681	CTGCAGAGCATATACACAGAGAAGCCCTGCACAACACACAGCTTCAGCTTACTCTGTGGCCAGT	1740
D	b	1681	CTGCAGAGCATATACACAGAGAAGCCCTGCACAACACACAGCTTACTCTGTGGCCAGT	1740
O	y	1741	GCCATACCAGCAGGAGCCTGTACTTGCGGCTCTTCTGCCCCGGAGGCTCTATTCAAGCAG	1800
D	b	1741	GCCATACCAGCAGGAGCCTGTACTTGCGGCTCTTCTGCCCCGGAGGCTCTATTCAAGCAG	1800
O	y	1801	ATCCAGGTGAAGCAGAACATCTCGGTGACCCCTTGCGACCTTTGGCCCAGCTTCCACAATA	1860
D	b	1801	ATCCAGGTGAAGCAGAACATCTCGGTGACCCCTTGCGACCTTTGGCCCAGCTTCCACAATA	1860
O	y	1861	GAGGCTCTCAGGACGAGGTGTGTGACGCTGTCTTTATACCTTAATTTCAAAGAGAAAGCGCTT	1920
D	b	1861	GAGGCTCTCAGGACGAGGTGTGTGACGCTGTCTTTATACCTTAATTTCAAAGAGAAAGCGCTT	1920
O	y	1921	TTTCAGGTGACCCCTGTGACACAAAAAGCAAGGTCTACCTGAGAGACCCCAACTGSGACGG	1980
D	b	1921	TTTCAGGTGACCCCTGTGACACAAAAAGCAAGGTCTACCTGAGAGACCCCAACTGSGACGG	1980
O	y	1981	GGGCTGGCATCGCTCACCTGTGTGTCTGTGAAACATCAGCTGTGCCAGAGACCAGTAGTGGCC	2040
D	b	1981	GGGCTGGCATCGCTCACCTGTGTGTCTGTGAAACATCAGCTGTGCCAGAGACCAGTAGTGGCC	2040
O	y	2041	TGCGTGACTTTCTTTAAAGAGCGAGCGGCGGTGTCTGCCAGACAGGGCGCGCATTTCAAG	2100
D	b	2041	TGCGTGACTTTCTTTAAAGAGCGAGCGGCGGTGTCTGCCAGACAGGGCGCGCATTTCAAG	2100
O	y	2101	ATCATCCAGAGACAGCAGGACCCGGGCTGAGAGATCTTCAGCGCTGACGAGAGATGTGCTC	2160
D	b	2101	ATCATCCAGAGAGACAGCAGGACCCGGGCTGAGAGATCTTCAGCGCTGACGAGAGATGTGCTC	2160
O	y	2161	CCCAAGCCAAGCTTCCACCATTCACAGCTTTGTGGTCAACATCTCTAACTGCAAGCCCAAG	2220
D	b	2161	CCCAAGCCAAGCTTCCACCATTCACAGCTTTGTGGTCAACATCTCTAACTGCAAGCCCAAG	2220
O	y	2221	AGCGGCAAGCAGCTAGACCTGCTCTTCTCGGGGACACTTACCCCAAGAGACTGTGACATTTG	2280
D	b	2221	AGCGGCAAGCAGCTAGACCTGCTCTTCTCGGGGACACTTACCCCAAGAGACTGTGACATTTG	2280
O	y	2281	ACTGTCACTCTCANTCGCAGCGGTGGAGGTGGAGTCTTACTGTGTCTGCCCCCTGGGCTC	2340
D	b	2281	ACTGTCACTCTCANTCGCAGCGGTGGAGGTGGAGTCTTACTGTGTCTGCCCCCTGGGCTC	2340
O	y	2341	ATCATTTCTGTGTGTGAAAAAGAAAAGAAAGAACAAACAGGGCCGCTGTGGGTATC	2400
D	b	2341	ATCATTTCTGTGTGTGAAAAAGAAAAGAAAGAACAAACAGGGCCGCTGTGGGTATC	2400
O	y	2401	TACAATGGCAACATCAATACAGATGAGATGCGGAGGCGAGCCAAAAAAGTTTTACAGAAAAGGGA	2460
D	b	2401	TACAATGGCAACATCAATACAGATGAGATGCGGAGGCGAGCCAAAAAAGTTTTACAGAAAAGGGA	2460
O	y	2461	AAGGCAATGATCGCATGTGTATGCACTCATCGAGACACCATTGTAATATGGGCACTGTG	2520
D	b	2461	AAGGCAATGATCGCATGTGTATGCACTCATCGAGACACCATTGTAATATGGGCACTGTG	2520
O	y	2521	CTACAGGATTTCCAGCGGCTCTTCTCTGCAAGCAGAGGTGAGACCTTACCGGCGCTTCCAG	2580
D	b	2521	CTACAGGATTTCCAGCGGCTCTTCTCTGCAAGCAGAGGTGAGACCTTACCGGCGCTTCCAG	2580
O	y	2581	GGGACCAATGGGGGTGTGTCTGTCCGCCACGACCAACATATGCTCCAGGGCCCAACTGTCA	2640
D	b	2581	GGGACCAATGGGGGTGTGTCTGTCCGCCACGACCAACATATGCTCCAGGGCCCAACTGTCA	2640
O	y	2641	AAGTTGGCCACTGAGAGACCACTCTCTGCCCTCCTCTGAGTGTGAAGTGAACCGTAC	2700
D	b	2641	AAGTTGGCCACTGAGAGACCACTCTCTGCCCTCCTCTGAGTGTGAAGTGAACCGTAC	2700
O	y	2701	ACCTTCTCCCATCCCAACAATGGGATGTAAAGCAGCAAGAGACAGACATTCCTTACTG	2760
D	b	2701	ACCTTCTCCCATCCCAACAATGGGATGTAAAGCAGCAAGAGACAGACATTCCTTACTG	2760
D	b	2701	ACCTTCTCCCATCCCAACAATGGGATGTAAAGCAGCAAGAGACAGACATTCCTTACTG	2760
O	y	2761	AACACTCAGAGACCCATGGAGGACAGACAGAAATTAATTCATTCACAGCGCTTTGCTGA	2820
D	b	2761	AACACTCAGAGACCCATGGAGGACAGACAGAAATTAATTCATTCACAGCGCTTTGCTGA	2820
O	y	2821	GTTTATATAAGCAGGGCACTGAGACACCCGTCGCTGTCTTAACCAAGAAATCTTAAGAA	2880
D	b	2821	GTTTATATAAGCAGGGCACTGAGACACCCGTCGCTGTCTTAACCAAGAAATCTTAAGAA	2880
O	y	2881	GAGGAATTAATACAGAAAGAAACGACGAGGAGTTTCTGTGACACCGCCAACTTACATTTGC	2940
D	b	2881	GAGGAATTAATACAGAAAGAAACGACGAGGAGTTTCTGTGACACCGCCAACTTACATTTGC	2940
O	y	2941	TCAGTGAATCATTTCTAAGGGGCAAGACATTTGAAATGATGAATTCCAATCTGATACAGT	3000
D	b	2941	TCAGTGAATCATTTCTAAGGGGCAAGACATTTGAAATGATGAATTCCAATCTGATACAGT	3000
O	y	3001	CATGACAGCTCATGTGTCTCTCACTTAAGGCTGTGCGGTTAGCCAGCCTGTATATGAGAG	3060
D	b	3001	CATGACAGCTCATGTGTCTCTCACTTAAGGCTGTGCGGTTAGCCAGCCTGTATATGAGAG	3060
O	y	3061	AGAGAGGCGTGAACATCAGATATGAGGTGAGAGAACCCGAGTTCAGAGGTCTTAAC	3120
D	b	3061	AGAGAGGCGTGAACATCAGATATGAGGTGAGAGAACCCGAGTTCAGAGGTCTTAAC	3120
O	y	3121	AGAGGCTTGCCCTCTTCAGAGCAACAGTTCCAATTCGAAGAGCCTTACCTGAGTCCCTTA	3180
D	b	3121	AGAGGCTTGCCCTCTTCAGAGCAACAGTTCCAATTCGAAGAGCCTTACCTGAGTCCCTTA	3180
O	y	3181	CTCTCAGTGGGGTCCCAAGAGATGAAGACAAATGTGCTTTTATATTAATTTATTTGG	3240
D	b	3181	CTCTCAGTGGGGTCCCAAGAGATGAAGACAAATGTGCTTTTATATTAATTTATTTGG	3240
O	y	3241	TGCTCCTGTGTATTTAAGAGATCAAAATGTAATTAACACCTAGCTTTTACACCTGACTTA	3300
D	b	3241	TGCTCCTGTGTATTTAAGAGATCAAAATGTAATTAACACCTAGCTTTTACACCTGACTTA	3300
O	y	3301	GTAATTAATCATACTAATCTGTTGTGATGCGGTGTGTGACTTCTACTGACCGCTAGAT	3360
D	b	3301	GTAATTAATCATACTAATCTGTTGTGATGCGGTGTGTGACTTCTACTGACCGCTAGAT	3360
O	y	3361	AAAGCTGTGCTGTCCCGGAGGTGGGATTAATTTACAACTGTGTCCACACAGAAAAGA	3420
D	b	3361	AAAGCTGTGCTGTCCCGGAGGTGGGATTAATTTACAACTGTGTCCACACAGAAAAGA	3420

QY 3841 GTTCTTCCATTTTCAGCTTCTAGAGGCTGGCTGCATTTCCAGGCTCCAGTGGCTC 3900
DB 3841 GTTCTTCCATTTTCAGCTTCTAGAGGCTGGCTGCATTTCCAGGCTCCAGTGGCTC 3900
QY 3901 AAGCTTTTCTCAGATGAGATCAGTGTGACACTGGCCCTCCCACTTCTCTTGTGACTAC 3960
DB 3901 AAGCTTTTCTCAGATGAGATCAGTGTGACACTGGCCCTCCCACTTCTCTTGTGACTAC 3960
QY 3961 AAAGCCACAGAGAGATCCAGGATATCTCTCCATCTAAAGATCCTTCAATCATCTGGA 4020
DB 3961 AAAGCCACAGAGAGATCCAGGATATCTCTCCATCTAAAGATCCTTCAATCATCTGGA 4020
QY 4021 AGAGCTTTTGCATGCAAGACAAACATAGCCAGGTGGGATTAAGACAGGACATCTT 4080
DB 4021 AGAGCTTTTGCATGCAAGACAAACATAGCCAGGTGGGATTAAGACAGGACATCTT 4080
QY 4081 TGGGCTGTCTTATTTCTCCCTACACACACTTCTCCACTGACTACTCTCCACAGAGAGGCTA 4140
DB 4081 TGGGCTGTCTTATTTCTCCCTACACACACTTCTCCACTGACTACTCTCCACAGAGAGGCTA 4140
QY 4141 CAAATATGATCTGGGACACAGGAGATGTTTGTAGCTGGGACTCTAACTTAAAAA 4200
DB 4141 CAAATATGATCTGGGACACAGGAGATGTTTGTAGCTGGGACTCTAACTTAAAAA 4200
QY 4201 ACCCCAGATCAGAGATCTGGCCATGCTGGGGCTCACATTCCTACCTAGCAACAATGGC 4260
DB 4201 ACCCCAGATCAGAGATCTGGCCATGCTGGGGCTCACATTCCTACCTAGCAACAATGGC 4260
QY 4261 TGGAGCTGGGACACAGCTCTGCTTTAGAGGGGCTCCACTCTCCAGGCTCACACAGC 4320
DB 4261 TGGAGCTGGGACACAGCTCTGCTTTAGAGGGGCTCCACTCTCCAGGCTCACACAGC 4320
QY 4321 CCACACTACGCTTATCTATCTCCCAATAGAGGCTAAGTGTGTTCTTCTACTGATCATG 4380
DB 4321 CCACACTACGCTTATCTATCTCCCAATAGAGGCTAAGTGTGTTCTTCTACTGATCATG 4380
QY 4381 CCCCTGCAAGTTGCTTATTTATGTAAGAAAAAGAAAGCTGGGATTAATCTCTAATCAGG 4440
DB 4381 CCCCTGCAAGTTGCTTATTTATGTAAGAAAAAGAAAGCTGGGATTAATCTCTAATCAGG 4440
QY 4441 TGAGTAGACATGAGACCAATGTGTGCTACATTAACCTTTTCTTTTCTTTTCT 4500
DB 4441 TGAGTAGACATGAGACCAATGTGTGCTACATTAACCTTTTCTTTTCTTTTCTTTTCT 4500
QY 4501 TTTTCTTTTCTTTTAAATGAGACAGAGATCTCATTTCTGTGCTAGGCTGAGTGCAG 4560
DB 4501 TTTTCTTTTCTTTTAAATGAGACAGAGATCTCATTTCTGTGCTAGGCTGAGTGCAG 4560
QY 4561 TGGGCAATCTCGGCTCACTGCAACCTTGGCTCTGAGGCTCAGCAATTTCTCCACCTC 4620
DB 4561 TGGGCAATCTCGGCTCACTGCAACCTTGGCTCTGAGGCTCAGCAATTTCTCCACCTC 4620
QY 4621 AGCCCTCCAAATAGCTGGGATCACTGGGCAAAACCAACCATGCCGCAATTTTGTATTT 4680
DB 4621 AGCCCTCCAAATAGCTGGGATCACTGGGCAAAACCAACCATGCCGCAATTTTGTATTT 4680
QY 4681 TTTGTAGAGACAGGTTTACCATGTTGCCAGGCTGGTCTCAACCTCTGGGCTCAAGC 4740
DB 4681 TTTGTAGAGACAGGTTTACCATGTTGCCAGGCTGGTCTCAACCTCTGGGCTCAAGC 4740
QY 4741 AATCTCTGCTGGCTGCCCTCCCAAAAGTCTGGGATTAAGATGTGAGCCACCGCATCCAG 4800
DB 4741 AATCTCTGCTGGCTGCCCTCCCAAAAGTCTGGGATTAAGATGTGAGCCACCGCATCCAG 4800
QY 4801 CCCCAACCCCATTTATACCAATTAACCTGCCAGTACCTGAGCTTTGCTTCTCTAC 4860
DB 4801 CCCCAACCCCATTTATACCAATTAACCTGCCAGTACCTGAGCTTTGCTTCTCTAC 4860
QY 4861 CCTGCTCTGATCTGAGAGAGAGGATTAATGTTATAGCTTGCAGACATCCCAAGTT 4920
DB 4861 CCTGCTCTGATCTGAGAGAGAGGATTAATGTTATAGCTTGCAGACATCCCAAGTT 4920

QY 4921 CAATATTTCTGGGCAAAAACCTCTCTCAAAAAATTAATGTACTTATTTAATCAATGA 4980
DB 4921 CAATATTTCTGGGCAAAAACCTCTCTCAAAAAATTAATGTACTTATTTAATCAATGA 4980
QY 4981 ATTCACTTGGAAATGCAACCGGCTCAACTGTTCACATGGGCTAAATGAAGAAATTTTA 5040
DB 4981 ATTCACTTGGAAATGCAACCGGCTCAACTGTTCACATGGGCTAAATGAAGAAATTTTA 5040
QY 5041 TAGTCTCTTAATGAGCTGTACTGCAAGACCTCTTGAACACTTTTCCAGAGATAGATAT 5100
DB 5041 TAGTCTCTTAATGAGCTGTACTGCAAGACCTCTTGAACACTTTTCCAGAGATAGATAT 5100
QY 5101 TTAAGTCAATGCCCTTGGCTTGCATATGAGCACCTTTCCCTTGTGAAGTCTGTCTGC 5160
DB 5101 TTAAGTCAATGCCCTTGGCTTGCATATGAGCACCTTTCCCTTGTGAAGTCTGTCTGC 5160
QY 5161 CCAGTGACCTTGGCTTGTGAGCCGGAATGCTGACCTCTGCAATTAAGGGCCAAAGAGGG 5220
DB 5161 CCAGTGACCTTGGCTTGTGAGCCGGAATGCTGACCTCTGCAATTAAGGGCCAAAGAGGG 5220
QY 5221 CTGCGGCTTCCCTCCCTCACTGAAAGCCCTTATTTGAATTCATCTGTGAGCCCTAGC 5280
DB 5221 CTGCGGCTTCCCTCCCTCACTGAAAGCCCTTATTTGAATTCATCTGTGAGCCCTAGC 5280
QY 5281 CCTCATTTCTGACATTTCCCAACCTCCAGCCCTTCCAAAGACAGACTAGTGGCTGCG 5340
DB 5281 CCTCATTTCTGACATTTCCCAACCTCCAGCCCTTCCAAAGACAGACTAGTGGCTGCG 5340
QY 5341 ATTCCACCAAGTGGGATTTGGGCTTCCCTTGAAGGCTGCTGCTACCATATCCGACAT 5400
DB 5341 ATTCCACCAAGTGGGATTTGGGCTTCCCTTGAAGGCTGCTGCTACCATATCCGACAT 5400
QY 5401 CACTGTGCTCTGCAAGACACACAGCTGGCCATTTTCTTCAACTGAGGCTCAAAAACCTC 5460
DB 5401 CACTGTGCTCTGCAAGACACACAGCTGGCCATTTTCTTCAACTGAGGCTCAAAAACCTC 5460
QY 5461 TGGACAAATGCTGGCTCTGAGACACATATTTCTGAGGCTGCTCCACAGAGAGGGG 5520
DB 5461 TGGACAAATGCTGGCTCTGAGACACATATTTCTGAGGCTGCTCCACAGAGAGGGG 5520
QY 5521 CCAGGCTGAGGAACCTGGCTCTTCTTAAAGCCAGGCCCCACTTACATTAACAACTT 5580
DB 5521 CCAGGCTGAGGAACCTGGCTCTTCTTAAAGCCAGGCCCCACTTACATTAACAACTT 5580
QY 5581 TCAGGCTCACTGGAACAGTGAAGTGCATTTGTTGAAGCCCTACTGATGCCAGCCACT 5640
DB 5581 TCAGGCTCACTGGAACAGTGAAGTGCATTTGTTGAAGCCCTACTGATGCCAGCCACT 5640
QY 5641 GCTCATCCAGTGTGCTGCAATGCTTACGAGGAAGCCAGGCAATGCAAGACTGGTCTCT 5700
DB 5641 GCTCATCCAGTGTGCTGCAATGCTTACGAGGAAGCCAGGCAATGCAAGACTGGTCTCT 5700
QY 5701 AATGCTGTGCTATGCAAGAGGGAAGGCTCAAGGAAGTCAACTGGGCAAGCA 5760
DB 5701 AATGCTGTGCTATGCAAGAGGGAAGGCTCAAGGAAGTCAACTGGGCAAGCA 5760
QY 5761 CAAGCCACCGGACATGAGCTTGTGAAGGTTAGCAGACGTGGTGTGATCTGCACT 5820
DB 5761 CAAGCCACCGGACATGAGCTTGTGAAGGTTAGCAGACGTGGTGTGATCTGCACT 5820
QY 5821 GCTTCACTGGAATTAATTTATTTCAATGCAATACTTTTGTAGTGCAATTTATTCATTT 5880
DB 5821 GCTTCACTGGAATTAATTTATTTCAATGCAATACTTTTGTAGTGCAATTTATTCATTT 5880
QY 5881 CTGTGCTTAAATTAACAAATGTACCAAAAAAACAAGTATCAAGCTTTAAGTGTCTGG 5940
DB 5881 CTGTGCTTAAATTAACAAATGTACCAAAAAAACAAGTATCAAGCTTTAAGTGTCTGG 5940
QY 5941 CTACTTGTCCCTGGTTGAGTGAAGGCCCGGTTTCCAGTGTGTTGACTGTGACAGGCTC 6000
DB 5941 CTACTTGTCCCTGGTTGAGTGAAGGCCCGGTTTCCAGTGTGTTGACTGTGACAGGCTC 6000
QY 6001 AGCATGGGCTCAGCAGATGTGTCTTAAATTTGTGATGATACAGAAACCAAGCTTGGG 6060

|||||
Db 6001 AGATGGCTCAACAGATGCTGCTAATTTGGATGATACGAAGACGAGCTTGG 6060
Qy 6061 ATACAAATCTCTTCTCTTCAATTTGATGCCGTGCTGATGAGACAGATGTTTTGTC 6120
|||||
Db 6061 ATACAAATCTCTTCTCTTCAATTTGATGCCGTGCTGATGAGACAGATGTTTTGTC 6120
Qy 6121 GGAATATTAATTAATGCTGAGCTGCGCCAAAAA 6163
|||||
Db 6121 GGAATATTAATTAATGCTGAGCTGCGCCAAAAA 6163
RESULT 2
AY026461 5963 bp mRNA linear pri 17-AUG-2001
LOCUS AY026461
DEFINITION Homo sapiens CUB domain containing protein 1 (CDCP1) mRNA, complete cds.
ACCESSION AY026461
VERSION AY026461.1 GI:14328878
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5963)
Schertl-Mostagener, M., Sommergruber, W., Abseher, R., Hauptmann, R., Ambros, P. and Schmeitler, N.
Identification of a novel gene, CDCP1, overexpressed in human colorectal cancer
Oncogene 20 (32), 4402-4408 (2001)
JOURNAL 11466521
MEDLINE 21359860
PUBMED 2 (bases 1 to 5963)
REFERENCE 2 (bases 1 to 5963)
Schertl-Mostagener, M., Sommergruber, W., Abseher, R., Hauptmann, R., Ambros, P. F. and Schmeitler, N.
Direct Submision
Submitted (31-JAN-2001) Exploratory Research, Boehringer Ingelheim, Austria, Dr. Boehringerasse 5-11, Vienna 1121, Austria
FEATURES
source
1..5963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p23-p21"
/cell_line="A549; Calu 6"
1..5963
/gene="CDCP1"
1..82
/gene="CDCP1"
83..2593
/gene="CDCP1"
/note="transmembrane protein"
/codon_start=1
/product="CUB domain containing protein 1"
/protein_id="AAK02058.1"
/db_xref="GI:14328879"
/translation="MAGLNCGVSTALLGVLLGAARLPFGAFAETALPRESNTYILI
KLGPTPLAKPCYIVISKRHITMISIKSERIVFPGSCSPNNHVIETQKIDCMG
PCPGEVLDLPSTSLPTLNTFINDVKHAKISFISGLSEFIRLPRIIGEGSPDVT
HSISGRIDATVIRIGTSCNGTVSRKMQEGYKMLHLHLPFRNPGVSGFISNRSIK
RLCIIIEVEEGSATLMSANPEGPDELTMTQFVPAHRAVSGLFNLSNER
KEEREYXIRGSTTPREVKLEDKOPGNACGNFNLSLQCDODASPGILRIQFOYLV
OHFONESKTIYVDLSNERAKSLTEPRVKOSKRFPECFVCLSRCSNLTLSG
SKHKISFLCDDTLRLMNVKTIISCTDHYCQRKSYSLQVPSDILHLPVELHDFSKL
LVKDRSLVLPFAQKLODHKEPCNTSFYSLVASLPSODLYGSGFPGSISIKQI
VKONISVLRFPAPFOEASRQGLTVSPFYKEEGVETVPDRSKYLYLTPMDR
GLPSLTSVMNITSVPDOYACLIFFKERSGVVCGTGRAFMIIQEDRTAELEISLSD
VLKPSFHHHSFVNISSCSPTSGKQDLLEVTLPRLVDLTILIAVGGVLLS
ALGLIICVKKKKKKTKNGPAVGIINGNTMPQPKFQGRKDNDSHVAIVEDT
NVYGLLDSSGSLQPEVDYRPFQGTMGVCPRPPTICSAAPAKLATEBPPRSP
PESESEPTTFSSHPNNGDVSSKTDIPILTLNTOPEMPAE"
2594..5963
/gene="CDCP1"

BASE COUNT 1463 a 1634 c 1392 g 1474 t
ORIGIN
Query Match 96.8%; Score 5963; DB 9; Length 5963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 201 CGAGAGTGGCCGGGCTTGGGGCCGCTGGGGCCCGCTCCCAACCGCTTCCACC 260
Db 1 CGAGAGTGGCCGGGCTTGGGGCCGCTGGGGCCCGCTCCCAACCGCTTCCACC 60
Qy 261 GAGCGGAGGCGTCCGAGATCATGGCGGCGCTGAACCTGGGGGTCTATTCGACATGCT 320
Db 61 GAGCGGAGGCGTCCGAGATCATGGCGGCGCTGAACCTGGGGGTCTATTCGACATGCT 120
Qy 321 AGGGGTTCTGCTGCTGGGTCGGCGCGCTGCCGCGGGGAGAAAGCTTTGAGATTGC 380
Db 121 AGGGGTTCTGCTGCTGGGTCGGCGCGCTGCCGCGGGGAGAAAGCTTTGAGATTGC 180
Qy 381 TCTGCCACGAGAAAGCAACATTCATGTTCTATAAGCTGGGACCCGACTGCTGGC 440
Db 181 TCTGCCACGAGAAAGCAACATTCATGTTCTATAAGCTGGGACCCGACTGCTGGC 240
Qy 441 AAAACCTGTACATGCTCATTTCTAAAGACATTAACCATGTGTCATCAAGTCTGG 500
Db 241 AAAACCTGTACATGCTCATTTCTAAAGACATTAACCATGTGTCATCAAGTCTGG 300
Qy 501 AGAAGATATGCTTACCTTACCTTAGCTCCAGAGTCTGAGAAATCACTTTGTATAGAT 560
Db 301 AGAAGATATGCTTACCTTACCTTAGCTCCAGAGTCTGAGAAATCACTTTGTATAGAT 360
Qy 561 CCAAGAAATATATGACGTATGTCAGGCCCATGTCTTTGGGAGAGTTACGTTACGCC 620
Db 361 CCAAGAAATATATGACGTATGTCAGGCCCATGTCTTTGGGAGAGTTACGTTACGCC 420
Qy 621 CTCGACATCGTTGTTGCCCTCACTCAACAGAACTTTCATCTGGAGTCAAACTCATAA 680
Db 421 CTCGACATCGTTGTTGCCCTCACTCAACAGAACTTTCATCTGGAGTCAAACTCATAA 480
Qy 681 GAGCATCGTTTAGAGCTGACGTTTTCATCCCTGCTGAGGACATCGGTCGGGTGA 740
Db 481 GAGCATCGTTTAGAGCTGACGTTTTCATCCCTGCTGAGGACATCGGTCGGGTGA 540
Qy 741 GAGCTGCCAGAGGAGTCACTCACTCAGGCGGCAATGATGCCAGCGGTGAG 800
Db 541 GAGCTGCCAGAGGAGTCACTCACTCAGGCGGCAATGATGCCAGCGGTGAG 600
Qy 801 GATCGAACTTGTGACAGCAATGACATGTGTCGGGATCAAGTCAAGAAAGAGTGA 860
Db 601 GATCGAACTTGTGACAGCAATGACATGTGTCGGGATCAAGTCAAGAAAGAGTGA 660
Qy 861 AATGGCTTACACTCCCATGTTCCACCCAGAAATGTCTCGGGCTTCAGCATTCGAAA 920
Db 661 AATGGCTTACACTCCCATGTTCCACCCAGAAATGTCTCGGGCTTCAGCATTCGAAA 720
Qy 921 CCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTGTTGAGGGTAAAGGCTAGC 980
Db 721 CCGCTCATCTATAAAGCTGTGTCATCATGAGTCTGTGTTGAGGGTAAAGGCTAGC 780
Qy 981 AACCTATATCTGCGCAACTACCCAGAAAGCTTCCCGAGAGTATGATGACGTGGCA 1040
Db 781 AACCTATATCTGCGCAACTACCCAGAAAGCTTCCCGAGAGTATGATGACGTGGCA 840
Qy 1041 GTTGTGCTTCTGACACACTGGGGGAGCGTCTCTCTCAACTTCAACTCTTCAA 1100
Db 841 GTTGTGCTTCTGACACACTGGGGGAGCGTCTCTCTCAACTTCAACTCTTCAA 900
Qy 1101 CTGTGAGAGAGAGAGAGGAGGTTGTAATCTATATCCGGGCTCCACCAACCCCGA 1160
Db 901 CTGTGAGAGAGAGAGAGGAGGTTGTAATCTATATCCGGGCTCCACCAACCCCGA 960
Qy 1161 GGTGTCAAGCTGGAGAGCAAGCAGCTGGGAACATGGGGAACCTTCAACTCTCTCT 1220
Db 961 GGTGTCAAGCTGGAGAGCAAGCAGCTGGGAACATGGGGAACCTTCAACTCTCTCT 1220

D 961 GGTGTTCAAGTGGAGCAAGCAAGCCTGGGAACATGGCGGGGAATTCAAACCTCTCTCT 1020
Q 1221 GCAAGGCTGTGACCAAGATGCCCAAGTCCAGGAGTCTCCGGCTGCAGTTCCAAGTTT 1280
D 1221 GCAAGGCTGTGACCAAGATGCCCAAGTCCAGGAGTCTCCGGCTGCAGTTCCAAGTTT 1080
Q 1281 GGTCCACATCCACAAAATGAAGCATAAATCTACGTGGTTGACTTGAATTAAGCG 1340
D 1081 GGTCCACATCCACAAAATGAAGCATAAATCTACGTGGTTGACTTGAATTAAGCG 1140
Q 1341 AGCCATGCTACATCAATCGAGCCGCGTCAACAGCGCGCAAGTTGTCCTG 1400
D 1141 AGCCATGCTACATCAATCGAGCCGCGTCAACAGCGCGCAAGTTGTCCTG 1200
Q 1401 CTGTTGCTGTGTAGAAATCTCGACCTGCAGTACCACTCAACCTGCATCTGGCTC 1460
D 1201 CTGTTGCTGTGTAGAAATCTCGACCTGCAGTACCACTCAACCTGCATCTGGCTC 1260
Q 1461 CAACACAAAATCTCTCTCTCTTGTGATGATCTGACACGTCTGTGATGAATGTGAAA 1520
D 1261 CAACACAAAATCTCTCTCTCTTGTGATGATCTGACACGTCTGTGATGAATGTGAAA 1320
Q 1521 AACCATAGCTGACAGACCAAGCAGTACTGCAAAAGGAATCTACTACTCCAGTGGC 1580
D 1321 AACCATAGCTGACAGACCAAGCAGTACTGCAAAAGGAATCTACTACTCCAGTGGC 1380
Q 1581 CAGTGACATCTCCACCTGCTGTGAGCTGATGATCTCTGGAAGCTGTGTGCC 1640
D 1381 CAGTGACATCTCCACCTGCTGTGAGCTGATGATCTCTGGAAGCTGTGTGCC 1440
Q 1641 CAAGGACAGGCTGACCTGCTGTGAGCTGATGATCTCTGGAAGCTGTGTGCC 1700
D 1441 CAAGGACAGGCTGACCTGCTGTGAGCTGATGATCTCTGGAAGCTGTGTGCC 1500
Q 1701 GAAGCCTGCAACACAGCTTCACTACTGCTGACCTGCTACTACCAAGCAGGACT 1760
D 1501 GAAGCCTGCAACACAGCTTCACTACTGCTGACCTGCTACTACCAAGCAGGACT 1560
Q 1761 GTACTTGGCTCTCTTGGCCGGAGGCTCTATCAAGCAGATCCAGGTGAAGCAACAT 1820
D 1561 GTACTTGGCTCTCTTGGCCGGAGGCTCTATCAAGCAGATCCAGGTGAAGCAACAT 1620
Q 1821 CTCGGTACCTTGGACCTTGGCCAGCTTCCAAACAGAGGCTTCCAGGAGGCT 1880
D 1621 CTCGGTACCTTGGACCTTGGCCAGCTTCCAAACAGAGGCTTCCAGGAGGCT 1680
Q 1881 GACGCTGCTTATACCTTATTCAAAGAGAGGCTTTCACGGTGACCCCTGACAC 1940
D 1681 GACGCTGCTTATACCTTATTCAAAGAGAGGCTTTCACGGTGACCCCTGACAC 1740
Q 1941 AAAAAGCAAGCTTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCCCTCACCTC 2000
D 1741 AAAAAGCAAGCTTACCTGAGAGACCCCAACTGGGACCGGGGCTGCCATCCCTCACCTC 1800
Q 2001 TGTGCTCTGGAACATCAAGCCGCGGACAGACAGTGGCTGCTGACTTCTTTAAGA 2060
D 1801 TGTGCTCTGGAACATCAAGCCGCGGACAGACAGTGGCTGCTGACTTCTTTAAGA 1860
Q 2061 GCGAGGCGGCTGTGCTGCGAGACAGGCGCGCATTCATGATTCATCCAGAGACGCGAC 2120
D 1861 GCGAGGCGGCTGTGCTGCGAGACAGGCGCGCATTCATGATTCATCCAGAGACGCGAC 1920
Q 2121 CCGGGCTGAGAGATCTTACCTGAGAGAGATGTGCTCCCAAGCCAAAGCTTCCACA 2180
D 1921 CCGGGCTGAGAGATCTTACCTGAGAGAGATGTGCTCCCAAGCCAAAGCTTCCACA 1980
Q 2181 TCACAGCTTCTGGGTCAACATCTTAACCTGAGAGCCCAAGCGGCAAGCTAGACT 2240
D 1981 TCACAGCTTCTGGGTCAACATCTTAACCTGAGAGCCCAAGCGGCAAGCTAGACT 2040
Q 2241 GCTCTTCTGGGTGACCTTACCCCAAGAGCTGTGAGCTTGAATCTCATCTGAGAGC 2300
D 2041 GCTCTTCTGGGTGACCTTACCCCAAGAGCTGTGAGCTTGAATCTCATCTGAGAGC 2100

Q 2301 GGTGGAGGTGAGCTTACTGCTGTGCTGCCCTGGGCTCATCTATTTGCTGTGAAAAA 2360
D 2101 GGTGGAGGTGAGCTTACTGCTGTGCTGCCCTGGGCTCATCTATTTGCTGTGAAAAA 2160
Q 2361 GAAGAAAAAGAACAAACAAAGGCGCGCTGTGGGTATCTCAATGCAACATCAATAC 2420
D 2161 GAAGAAAAAGAACAAACAAAGGCGCGCTGTGGGTATCTCAATGCAACATCAATAC 2220
Q 2421 TGAGATGCCGAGGACGCCAAAAAGTTTCAGAAAAGGCGGAAGCAATGACTCCATGT 2480
D 2221 TGAGATGCCGAGGACGCCAAAAAGTTTCAGAAAAGGCGGAAGCAATGACTCCATGT 2280
Q 2481 GTATGAGTCAATGAGACACACATGATATATGAGCATCTGTACAGATTCAGAGGCTC 2540
D 2281 GTATGAGTCAATGAGACACACATGATATATGAGCATCTGTACAGATTCAGAGGCTC 2340
Q 2541 CTTCCTGAGCAAGAGTGGACACCTTACCGGCGCTTCCAGGCGACATGGGGCTGTCC 2600
D 2341 CTTCCTGAGCAAGAGTGGACACCTTACCGGCGCTTCCAGGCGACATGGGGCTGTCC 2400
Q 2601 TCCCTCCCAACCAACATATGCTCCAGGCGCCCAACTCAAAAGTTGGCCTGAGAGAGC 2660
D 2401 TCCCTCCCAACCAACATATGCTCCAGGCGCCCAACTCAAAAGTTGGCCTGAGAGAGC 2460
Q 2661 ACTCTCTGCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCCACAA 2720
D 2461 ACTCTCTGCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCCACAA 2520
Q 2721 TGGGATGTTAGCAGCAAGAGACACAGATCCCTTACTGAAACCTCGAGGCGCATGGA 2780
D 2521 TGGGATGTTAGCAGCAAGAGACACAGATCCCTTACTGAAACCTCGAGGCGCATGGA 2580
Q 2781 GCCAGCAAAATTAATCTGATTCATTCAGAGCCTTGTGAGTTTCATTAACAGAGGACT 2840
D 2581 GCCAGCAAAATTAATCTGATTCATTCAGAGCCTTGTGAGTTTCATTAACAGAGGACT 2640
Q 2841 GAGACACCCGCTCGCTGTCTTACACAGAAATCTTAAGAGAGAGATTAACAGAGAA 2900
D 2641 GAGACACCCGCTCGCTGTCTTACACAGAAATCTTAAGAGAGAGATTAACAGAGAA 2700
Q 2901 CAGGAGAGGTTTCCGAGACACCGCCCACTTCAATGCTGCTAGTGGCTCATCTAAG 2960
D 2701 CAGGAGAGGTTTCCGAGACACCGCCCACTTCAATGCTGCTAGTGGCTCATCTAAG 2760
Q 2961 GCAAGACATTAATAATGATGATTCATCAATCGATACAGTATGACACTCATGTGCTC 3020
D 2761 GCAAGACATTAATAATGATGATTCATCAATCGATACAGTATGACACTCATGTGCTC 2820
Q 3021 TCAACTTAGGCTGTGCGTTAGCCAGCCTGTAAAGAGAGAGAGGCTGATCACCTA 3080
D 2821 TCAACTTAGGCTGTGCGTTAGCCAGCCTGTAAAGAGAGAGAGGCTGATCACCTA 2880
Q 3081 GCATAGGTTGACAGAGCCCTGATTCAGAGTATTAACAGAGGCTGCCCTCTCAGG 3140
D 2881 GCATAGGTTGACAGAGCCCTGATTCAGAGTATTAACAGAGGCTGCCCTCTCAGG 2940
Q 3141 ACAACAGTTCCAATTCAGAGAGGCTTACCTGAGTCCCTACTCTACTGGGGTCCCAAG 3200
D 2941 ACAACAGTTCCAATTCAGAGAGGCTTACCTGAGTCCCTACTCTACTGGGGTCCCAAG 3000
Q 3201 ATGAAAAAGCAATGTGCTTATTAATTAATTAATTAATTAATTAATTAATTAAGA 3260
D 3061 ATGAAAAAGCAATGTGCTTATTAATTAATTAATTAATTAATTAATTAATTAAGA 3060
Q 3261 GATCAAAATGATTAACCAACATGCTTTTACCTGACTTATTAATTAATTAATTAATG 3320
D 3061 GATCAAAATGATTAACCAACATGCTTTTACCTGACTTATTAATTAATTAATTAATG 3120
Q 3321 GTTGGATGCTGGGTTGTGACTTCTACTGACCCCTAGATTAACGCTGTGCTCCCA 3380
D 3121 GTTGGATGCTGGGTTGTGACTTCTACTGACCCCTAGATTAACGCTGTGCTCCCA 3180

OY	3381	GCGGTGGGAAATTAATTACATCTGTCCAAACCAGAAAAGAATGTGTGTGTTGACACA	3440
Db	3181	GGTGGTGGAATAATTACTAATCTGTCCAAACCAGAAAAGAATGTGTGTGTTGACACA	3240
OY	3441	TTCACATATCTGCCTTGTGATTAAGACACTTCCGTGATCTGTGAGCGGTGGTGGTATC	3500
Db	3241	TTCGACATATCTGTGTTGATTAAGACACTTCCGTGATCTGTGAGCGGTGGTGGTATC	3300
OY	3501	CCATTTGTGAAATTCATCTTGAATCCCATTGTCTATAGCTCTAGCAATTAAGAAAAATT	3560
Db	3301	CCATTTGTGAAATTCATCTTGAATCCCATTGTCTATAGCTCTAGCAATTAAGAAAAATT	3360
OY	3681	CATTAGTTTTCTGTGTGCACGTGACCCCATTTACTTGTGTGATTTAAAAAACACATTAGC	3740
Db	3481	CATTAGTTTTCTGTGTGCACGTGACCCCATTTACTTGTGTGATTTAAAAAACACATTAGC	3540
OY	3741	TTATATGCTCTGGGGATCAGAAATTCCAAATTCGATGTCCCTGANTGAAAAATCAGGTGCA	3800
Db	3541	TTATATGCTCTGGGGATCAGAAATTCCAAATTCGATGTCCCTGANTGAAAAATCAGGTGCA	3600
OY	3801	GCACAGCTGTGCTCTCTTCTGTAAGGCTCTGAGGGAGAAAGCGGTGCTGCCATTTCAAGCT	3860
Db	3601	GCACAGCTGTGCTCTCTTCTGTAAGGCTCTGAGGGAGAAAGCGGTGCTGCCATTTCAAGCT	3660
OY	3861	TCTAGAGGCTGGCTGCATTCCTCCAGGCTCCAGTGGCTGATGCAAGCTTTCTCACATGGCAT	3920
Db	3661	TCTAGAGGCTGGCTGCATTCCTCCAGGCTCCAGTGGCTGATGCAAGCTTTCTCACATGGCAT	3720
OY	3921	CACGTGTACACTGGCCCCCTCCACTTCCCTTGTGACTTACAAAGCCACACAGAAATGC	3980
Db	- 3721	CACGTGTACACTGGCCCCCTCCACTTCCCTTGTGACTTACAAAGCCACACAGAAATGC	3780
OY	3981	AGGATATCTCTCATCTATAAGATTCCTTATCATCTCTGTAAGGCTTTGCCATGTGAG	4040
Db	3781	AGGATATCTCTCATCTATAAGATTCCTTATCATCTCTGTAAGGCTTTGCCATGTGAG	3840
OY	4041	ACAACATAGCCACAGGTGGGATTTAGAGCACAGACATCTTTGGGTGCTGTTATTTGCC	4100
Db	3841	ACAACATAGCCACAGGTGGGATTTAGAGCACAGACATCTTTGGGTGCTGTTATTTGCC	3900
OY	4101	TACCACACCTTCTGTCACACTGTCCACAGAGAGGCTACAAAATGATCTGGCGCACAG	4160
Db	3901	TACCACACCTTCTGTCACACTGTCCACAGAGAGGCTACAAAATGATCTGGCGCACAG	3960
OY	4161	GGATGTTTTTTTAGCTTGGGACTTAACACTTAAAAAAACCCAGATCAGAAAGATCTG	4220
Db	3961	GGATGTTTTTTTAGCTTGGGACTTAACACTTAAAAAAACCCAGATCAGAAAGATCTG	4020
OY	4221	GCCATGCTGGGGCTCACATCTCFACCTAGCAACAACCTGGCTGGAGCCAGACAGCTCT	4280
Db	4021	GCCATGCTGGGGCTCACATCTCFACCTAGCAACAACCTGGCTGGAGCCAGACAGCTCT	4080
OY	4281	GCCCTTTAGAAAGGGGTGTCCACTTCACACAGTCCACACAGCCACACTAGCCCTATCACT	4340
Db	4081	GCCCTTTAGAAAGGGGTGTCCACTTCACACAGTCCACACAGCCACACTAGCCCTATCACT	4140
OY	4341	TCCCAACAATAGGCTTAAGTGTGTTGTTCTACTGATCAATGCCCTGCAGGTTGCATTTAT	4400
Db	4141	TCCCAACAATAGGCTTAAGTGTGTTGTTCTACTGATCAATGCCCTGCAGGTTGCATTTAT	4200
OY	4401	TGTAAATGAAAAAAGAAAGACGTGGGATTAATCTGTAAATCAGGTGAGTAGACATAGACCAA	4460
Db	4201	TGTAAATGAAAAAAGAAAGACGTGGGATTAATCTGTAAATCAGGTGAGTAGACATAGACCAA	4260
OY	4461	TGTGTGCTCAATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTATG	4520

Dd	4261	TGTGTGCTCACATYACCCCTTTTCCTTTTTTCTTTTTCTTTTTCTTTTTTCTTTTTTAAG	4320
Oy	4521	TGAGACAGATCTCATTCCTTGTCCTTAGCGTGAAGTGCAGTGGCCAACTCGGCTCACT	4580
Dd	43321	TGAGACAGATCTCATTCCTTGTCCTTAGCGTGAAGTGCAGTGGCCAACTCGGCTCACT	4380
Oy	4581	GCAACCTCTGCCCTCGGGCTCAGCAATYTCTCCACCCTCAAGCCTCCCAAATAGCTGGGA	4640
Dd	4381	GCAACCTCTGCCCTCGGGCTCAGCAATYTCTCCACCCTCAAGCCTCCCAAATAGCTGGGA	4440
Oy	4641	TGACTGGCAACAACCAACATGGCCAGCTAATTTTGTATTTTTGTAGAGACAGGTTTCA	4700
Dd	4441	TGACTGGCAACAACCAACATGGCCAGCTAATTTTGTATTTTTGTAGAGACAGGTTTCA	4500
Oy	4701	CCATGTGGCCAGGGTGGGTCTCAACCTCGTGGGCTCAAGCAATCTCGTGGCTGGCCCTC	4760
Dd	4501	CCATGTGGCCAGGGTGGGTCTCAACCTCGTGGGCTCAAGCAATCTCGTGGCTGGCCCTC	4560
Oy	4761	CCAAAGTGTGGGATTAACAGATGTAGGCCACCGCATCCAGCCCCAACCCCTCATTTATAC	4820
Dd	4561	CCAAAGTGTGGGATTAACAGATGTAGGCCACCGCATCCAGCCCCAACCCCTCATTTATAC	4620
Oy	4821	CAATTACCTGCCCACTAAGTGTGGACTTTTGTCTCTCACCCCTCTGTATCTGGAAG	4880
Dd	4621	CAATTACCTGCCCACTAAGTGTGGACTTTTGTCTCTCACCCCTCTGTATCTGGAAG	4680
Oy	4881	AGAGGATTATGTATPAGCTGTGCAGCAGTCCCAAGTCAATATTTCTGGGGCAAAA	4940
Dd	4681	AGAGGATTATGTATPAGCTGTGCAGCAGTCCCAAGTCAATATTTCTGGGGCAAAA	4740
Oy	4941	CTTCCCTTCAAAAAATAATGTAATCTTCAATTTGTAATGAATTCACCTTGGAAATGACCC	5000
Dd	4741	CTTCCCTTCAAAAAATAATGTAATCTTCAATTTGTAATGAATTCACCTTGGAAATGACCC	4800
Oy	5001	GCCTCAACTTGTTCACATGCGCATTAATGAAAGAAATTTATATGTCTCTAATAGCCGTGT	5060
Dd	4801	GCCTCAACTTGTTCACATGCGCATTAATGAAAGAAATTTATATGTCTCTAATAGCCGTGT	4860
Oy	5061	ACTGCAAGCCCTTTGAACACTTTCAGAGGATAGAGATATTTAAGTCAMGGCCCTTGGGCT	5120
Dd	4861	ACTGCAAGCCCTTTGAACACTTTCAGAGGATAGAGATATTTAAGTCAMGGCCCTTGGGCT	4920
Oy	5121	TGCCATATGGCACCTTTCCTTCTGAAAGCTGTGTCTCGCCCAAGTACCTTGGGCTTGT	5180
Dd	4921	TGCCATATGGCACCTTTCCTTCTGAAAGCTGTGTCTCGCCCAAGTACCTTGGGCTTGT	4980
Oy	5181	GAGCCGAGATGCTGACCCCTGCATTAAGAAGGCCCAAGAGAGGGCTCGGGCTTCCCTCTAC	5240
Dd	4981	GAGCCGAGATGCTGACCCCTGCATTAAGAAGGCCCAAGAGAGGGCTCGGGCTTCCCTCTAC	5040
Oy	5241	TGAABAAGCCCTTATTTGAAATTCACGTGTGTGAGAGCCCTTAGCCTCCATYCTGACATTTCCC	5300
Dd	5041	TGAABAAGCCCTTATTTGAAATTCACGTGTGTGAGAGCCCTTAGCCTCCATYCTGACATTTCCC	5100
Oy	5301	CACACTCCAGCCCTTCCCAAGCAGAGACTAGGTGGCCCTGCATTTCCACCCAAGGTGGATTT	5360
Dd	5101	CACACTCCAGCCCTTCCCAAGCAGAGACTAGGTGGCCCTGCATTTCCACCCAAGGTGGATTT	5160
Oy	5361	GGCCTTCTTAGGCTGGGCTACTTGTTCACATCACCAGCATCACTGTGTGCTGCAAGACA	5420
Dd	5161	GGCCTTCTTAGGCTGGGCTACTTGTTCACATCACCAGCATCACTGTGTGCTGCAAGACA	5220
Oy	5421	CCAGCTGGCCATTTTCTCTTCAACCTAAGGGCTTAAAACTCTGAGACAAATGTCTGGCTCCT	5480
Dd	5221	CCAGCTGGCCATTTTCTCTTCAACCTAAGGGCTTAAAACTCTGAGACAAATGTCTGGCTCCT	5280
Oy	5481	GAGACCAAGTATTTCTGTGAGCTGTGCTCACTAGTGAAGGGGCCAGGCTGAGAACCTCTGGC	5540
Dd	5281	GAGACCAAGTATTTCTGTGAGCTGTGCTCACTAGTGAAGGGGCCAGGCTGAGAACCTCTGGC	5340
Oy	5541	TCTTTTCTTTAAAGCCAGGGCCCACTTACATTAACAATTTCAAGGCTACTGGAACAGT	5600

Db 5341 TCTTTTCTTTAAAGCCAGGCCCACTTACATTAATACTTTCAGGCTCAGTGAAGAACT 5400
Qy 5601 GAAGTGCATTTTGTGAAGCCTTACTGCATGCCAGCCCACTGCTCATCCAGCTGGTCTGCC 5660
Db 5401 GAAGTGCATTTTGTGAAGCCTTACTGCATGCCAGCCCACTGCTCATCCAGCTGGTCTGCC 5460
Qy 5661 ATGCTTCAGAGGAAGGCCAGGCGCATGCGAGCTGCTCTATGCTGTGTGCTATTTGACA 5720
Db 5461 ATGCTTCAGAGGAAGGCCAGGCGCATGCGAGCTGCTCTATGCTGTGTGCTATTTGACA 5520
Qy 5721 GAAGGAAAGGCTCAAGGAAGAGTCAACTGGGCAAGCAAGCCCAAGCCGAGCATGGCC 5780
Db 5521 GAAGGAAAGGCTCAAGGAAGAGTCAACTGGGCAAGCAAGCCCAAGCCGAGCATGGCC 5580
Qy 5781 TTGGTAAAGGTTAGCAGAGCTGCTGTGTGATCTGAGTCTGAGTCTCAGTGAATAATTTA 5840
Db 5581 TTGGTAAAGGTTAGCAGAGCTGCTGTGTGATCTGAGTCTGAGTCTCAGTGAATAATTTA 5640
Qy 5841 TTCAATTGAGATTTTATTTAGTGGCATTTTATTCATTTCTGTGCTTTAATAACAA 5900
Db 5641 TTCAATTGAGATTTTATTTAGTGGCATTTTATTCATTTCTGTGCTTTAATAACAA 5700
Qy 5901 TGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTCCGCTTACTGTCCTGGTTCAG 5960
Db 5701 TGTACCAAAAAACAAGTATCAAGCTGTTTAAAGTCTTCCGCTTACTGTCCTGGTTCAG 5760
Qy 5961 TAGAGGCCCGCGTTCAGAGTGTGTGATCTGAGAGGCTCAGATGGGCTCAGCAGATGC 6020
Db 5761 TAGAGGCCCGCGTTCAGAGTGTGTGATCTGAGAGGCTCAGATGGGCTCAGCAGATGC 5820
Qy 6021 TGTCTTAAATTTGTGATGATACAGAAAGCCAGGCTTGGGATACAGTCTTCTCTTC 6080
Db 5821 TGTCTTAAATTTGTGATGATACAGAAAGCCAGGCTTGGGATACAGTCTTCTCTTC 5880
Qy 6081 ATTGATGCGCTGAGTCTGTGAGAGAGATGTTTTCGGAATAATAATAATAGTCT 6140
Db 5881 ATTGATGCGCTGAGTCTGTGAGAGAGATGTTTTCGGAATAATAATAATAGTCT 5940
Qy 6141 TGAAGTCTCGCCAAAAA 6163
Db 5941 TGAAGTCTCGCCAAAAA 5963

RESULT 3
AX353603
LOCUS AX353603 5897 bp DNA linear PAR 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0204508.
ACCESSION AX353603
VERSION AX353603.1 GI:18618676
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Schwaefel, N., Schell, M., Mostagaer, M., Sommergruber, W. and Abseher, R.
TITLE Tumour-associated antigen (b545), characterised by an amino acid
sequence as in seq. id. No. 4
JOURNAL Patent: WO 0204508-A 1 17-JAN-2002;
FEATURES
source location/Qualifiers
1..5897
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..214
215..2464
/note="unlabeled protein product"
/codon_start=1
/protein_id="CAD22953.1"
/db_xref="GI:18618676"
/translation="MLSIKSGRIIVTFSCSPENHVFIEIKNIDOMSGPCEGYO
LQSTSLPTLNTFLIMDYRAHKSIGLELOFSLPRLROIGPESGPDGYHSLSGRID
ATYVRIGTCSNCTVSRIRKQEGYKALHLPHPHKRVSVSFSTIANSSIRKLTIIISV
FEGESATILMSANYPGPFEDILTMQFVPAHLRASVSFLNLSNCEKREBERVEY

IPGSTNPEVFKIEDKOPGNMAGNFNLSLOGCDODASPCILRIQFOVLYOHPOBESN
KIIVYDLISNERAMSLTIERPVRKSRKFRVSCDLESRTCSNLTLTSGKRIISFL
CDLITRLMMNVEKTIISCTDHRKORRYSLOVSDILHLVEILDFEMKLLVPRDLS
LVYPAOKLQOHHEKPCNTFSYLSVAIPSDOLYSGSFQSGISIOIVKONISVT
LRFPAFPOEASROGLIVSFIRYFREGEVFTYPTDKSVYLRTPMDGLPSTSV
SMNISVPRDVAULTIFERKESGVCCGRAPMIIIOBRTABEELFSDIEVLPSPSH
HHSFVNIISGSPISGKODLILFVSLTPPTVDILVLIAGGVLLSLALGILIC
VKRRKKTNRKNGPVGILNGNINTEMPSQKSFREGEFTMTPMQOSSTPYMGICR
IPAPSCSQWPTPTGRSRAPWGSVLPHPYAPGPOLOSPLSLHLLPLSLRVNT
PSPILPTGM"
2465..5897
BASE COUNT 1464 a 1596 c 1364 g 1467 t 6 others
ORIGIN
Query Match 95.0%; Score 5854.2; DB 6; Length 5897;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5865; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
Qy 291 CCTGAACCTGGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGGTGGCGGCGCT 350
Db 25 CCTGAACCTGGGGGCTCTATCGCACTGCTAGGGGTTCTGCTGGTGGCGGCGCT 84
Qy 351 GCGCGGGGGGAGAAAGCTTTTGAATTTGCTCTGCGACAGAGAAAGCAATTAAGTCT 410
Db 85 GCGCGGGGGGAGAAAGCTTTTGAATTTGCTCTGCGACAGAGAAAGCAATTAAGTCT 144
Qy 411 CATAAAGCTGGGGAGCCCGACTGCTGGCAAAACCCGTATACATCGCATTTCTAAAG 470
Db 145 CATAAAGCTGGGGAGCCCGACTGCTGGCAAAACCCGTATACATCGCATTTCTAAAG 204
Qy 471 ACATATAACATGTTGTCATCAAGTCTGAGAAAGAAATAGTCTTTTACCTTTAGTGGCA 530
Db 205 ACATATAACATGTTGTCATCAAGTCTGAGAAAGAAATAGTCTTTTACCTTTAGTGGCA 264
Qy 531 GAGTCCGAGATACATTTGTCATAGAGATCCAGAAATATGACGTATGTCAGAGCC 590
Db 265 GAGTCCGAGATACATTTGTCATAGAGATCCAGAAATATGACGTATGTCAGAGCC 324
Qy 591 ATGTCCTTTTGGGAGGTTGAGTTCAGCCCTCGACATCGTTGTTGCTACCTTCACAG 650
Db 325 ATGTCCTTTTGGGAGGTTGAGTTCAGCCCTCGACATCGTTGTTGCTACCTTCACAG 384
Qy 651 AACTTTATCTGGGATGTCAAAGCTCATAGAGATGGGTTTAAAGTCTGAGTTTCCAT 710
Db 385 AACTTTATCTGGGATGTCAAAGCTCATAGAGATGGGTTTAAAGTCTGAGTTTCCAT 444
Qy 711 CCTTCGCTGAGGAGATCGGTCCGGGAGAGAGTCCGACAGAGAGTCACTGCTCAT 770
Db 445 CCTTCGCTGAGGAGATCGGTCCGGGAGAGAGTCCGACAGAGAGTCACTGCTCAT 504
Qy 771 CAGCGCCGAATCGATGCCACCGTGTGAGAGTGGAAACCTTTCGACCAATGACATGT 830
Db 505 CAGCGCCGAATCGATGCCACCGTGTGAGAGTGGAAACCTTTCGACCAATGACATGT 564
Qy 831 GTCCCGATCAAGATGCAAGAGAGTGAATAATGGCTTAACTCCCATGGTTCACCC 890
Db 565 GTCCCGATCAAGATGCAAGAGAGTGAATAATGGCTTAACTCCCATGGTTCACCC 624
Qy 891 CAGAAATGTCTCCGGCTTCAGCATTTGCAAAACCGCTCATTTAAAGCTGTGCATCAT 950
Db 625 CAGAAATGTCTCCGGCTTCAGCATTTGCAAAACCGCTCATTTAAAGCTGTGCATCAT 684
Qy 951 CGAGTCTGTGTTTGAAGGTAAGGCTCAGCAACCTGATGTCTGCAACTACCGAGAAG 1010
Db 685 CGAGTCTGTGTTTGAAGGTAAGGCTCAGCAACCTGATGTCTGCAACTACCGAGAAG 744
Qy 1011 CTTCCTGAGAGATGATCATGAGTGGAGATTGTGCTTCTGACACCTGCGGGCCAG 1070
Db 745 CTTCCTGAGAGATGATCATGAGTGGAGATTGTGCTTCTGACACCTGCGGGCCAG 804
Qy 1071 CGTCTCTTCTCTCACTTCAACCTCTCCAACTGTGAAGAGAGAGAGGCGGTTGAATA 1130
Db 805 CGTCTCTTCTCTCACTTCAACCTCTCCAACTGTGAAGAGAGAGAGGCGGTTGAATA 864

QY	1131	CTAATCCGGGGCTCCACACCCACCAACCCCGAGGTGTTCAAGCTGGAGGACAAAGCAGCTGG	1190
Db	865	CTACATCCCGGGGCTCCACACCCCAACCCCGAGGTGTTCAAGCTGGAGGACAAAGCAGCTGG	924
QY	1191	GAACATGGCGGGAAACTTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAAGTCC	1250
Db	925	GAACATGGCGGGAAACTTCAACCTCTCTGCAAGGCTGTGACCAAGATGCCAAAGTCC	984
QY	1251	AGGATCTCTCGGCTGCAGTTCCAAGTTTTGGTCCAAATCCACAATAATGAAAGCAATA	1310
Db	985	AGGATCTCTCGGCTGCAGTTCCAAGTTTTGGTCCAAATCCACAATAATGAAAGCAATA	1044
QY	1311	AATCTACGTGGTGTACTTGTAGTAATGAGCGAGCATGTCACTACACATCGAGCCACGGCC	1370
Db	1045	AATCTACGTGGTGTACTTGTAGTAATGAGCGAGCATGTCACTACACATCGAGCCACGGCC	1104
QY	1371	CGTCAAAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCGTGTGTCTAGAAATCTCGGACTG	1430
Db	1105	CGTCAAAACAGAGCCGCAAGTTTGTCCCTGGCTGTTTCGTGTGTCTAGAAATCTCGGACTG	1154
QY	1431	CAGTAGACACTCCCTGACATCTGGGTCCAAACACAAATCTCCTTGTGTAGTA	1490
Db	1165	CAGTAGACACTCCCTGACATCTGGGTCCAAACACAAATCTCCTTGTGTAGTA	1224
QY	1491	TCGTACACGTCTGTGTAGTAATGTGAAAAAAACATTAAGCTGCACAGACACACCGGTACTG	1550
Db	1225	TCGTACACGTCTGTGTAGTAATGTGAAAAAAACATTAAGCTGCACAGACACACCGGTACTG	1284
QY	1551	CCAAAGAAATCTCACTCACTCCAGGTGCCCAAGTGCATCTCCACCTGCTGTGAGCT	1610
Db	1285	CCAAAGAAATCTCACTCACTCCAGGTGCCCAAGTGCATCTCCACCTGCTGTGAGCT	1344
QY	1611	GCATGACTTTCCTCGGAAAGTGTGTGTGCCCAAGGACAGGTCACACTGCTGTGTGTGTC	1670
Db	1345	GCATGACTTTCCTCGGAAAGTGTGTGTGCCCAAGGACAGGTCACACTGCTGTGTGTGTC	1404
QY	1671	AGCCCAAGACTGGAGCAGCATACACAGGAAAGCCCTCAACACAGCTTCAGGTACTCT	1730
Db	1405	AGCCCAAGACTGGAGCAGCATACACAGGAAAGCCCTCAACACAGCTTCAGGTACTCT	1464
QY	1731	CGTGGCAATGACCATACCCAGCAGGAGACCTGTACTTGGCTCTTCTGCGCGGAGGCTC	1790
Db	1465	CGTGGCAATGACCATACCCAGCAGGAGACCTGTACTTGGCTCTTCTGCGCGGAGGCTC	1524
QY	1791	TATCAACCAATCCAGGTGAAGCAAAACATCTCGGTGACCTTGGCAGCTTTGGCCCCAG	1850
Db	1525	TATCAACCAATCCAGGTGAAGCAAAACATCTCGGTGACCTTGGCAGCTTTGGCCCCAG	1584
QY	1851	CTTCAACAAGAGGCTCCAGGACAGGAGGTCTGACGGTGTCTTTATTAACCTTAATTCAAAGA	1910
Db	1585	CTTCAACAAGAGGCTCCAGGACAGGAGGTCTGACGGTGTCTTTATTAACCTTAATTCAAAGA	1644
QY	1911	GGAAGGCGTTTTCACGGTGAACCCCTGACACAAAAAGCAAGGTCTACCTGAGGACCCCA	1970
Db	1645	GGAAGGCGTTTTCACGGTGAACCCCTGACACAAAAAGCAAGGTCTACCTGAGGACCCCA	1704
QY	1971	CTGGGACCGGGGCTCGCATCTCTGACCTCTGTGTCTTGTGAACATCAGCGTCCCAAGA	2030
Db	1705	CTGGGACCGGGGCTCGCATCTCTGACCTCTGTGTCTTGTGAACATCAGCGTCCCAAGA	1764
QY	2031	CCAGGTGGCTGCGCTGACTTCTTTTAAGGAGCGGACCGGCGTGGTGCACAGACAGGGCG	2090
Db	1765	CCAGGTGGCTGCGCTGACTTCTTTTAAGGAGCGGACCGGCGTGGTGCACAGACAGGGCG	1824
QY	2091	CGCATTCATGATCATCCAGGACGAGCGGACCCGGGTGAGGAGATCTTCAGCTTGGACGA	2150
Db	1825	CGCATTCATGATCATCCAGGACGAGCGGACCCGGGTGAGGAGATCTTCAGCTTGGACGA	1884
QY	2151	GGATGTGCTCCCAAGCAAGCTTCCACATCAGACTTCTGGTCCAAACATCTTAAGT	2210
Db	1885	GGATGTGCTCCCAAGCAAGCTTCCACATCAGACTTCTGGTCCAAACATCTTAAGT	1944

QY	2211	CAGCCCCACGAGCGCGCAAGACGCTAGACCTGCTCTTCTCGGTGACACATTACCCCAAGAC	2270
Db	1945	CAGCGCCACGAGCGCGCAAGACGCTAGACCGCTCTTCTCGGTGACACATTACCCCAAGAC	2004
QY	2271	TGTGGACTGTGACTCATCTCCATCCGACGGGGGGGAGGTGGAGGCTTACGTCTGTGCG	2330
Db	2005	TGTGGACTGTGACTCATCTCCATCCGACGGGGGGGAGGTGGAGGCTTACGTCTGTGCG	2054
QY	2331	CCTGGGCGCTCATCTTTTGTGCTGTGTGAAAAAGAAAGAAAAAGAACAAACAGGGCCCCG	2390
Db	2065	CCTGGGCGCTCATCTTTTGTGCTGTGTGAAAAAGAAAGAAAAAGAACAAACAGGGCCCCG	2124
QY	2391	TGTGGGTATCTACATAGGCACATCAATTACTGAGATCCGAGGACGCCAAAAAGTTTCA	2450
Db	2125	TGTGGGTATCTACATAGGCACATCAATTACTGAGATGCC - AGGCGACGCAAAAAAGTTTCA	2183
QY	2451	GAAGGGCGGAAGAGACAAATGACATCCCATGTGTATGACATCATCGAGACACCATGTGTATA	2510
Db	2184	GAAGGGCGGAAGAGACAAATGACATCCCATGTGTATGACATCATCGAGACACCATGTGTATA	2243
QY	2511	TGGGCAATCGCTAACAGAGATTCCAGACGGCTCCTTCTCGACGACAGAGTGGACACTACCG	2570
Db	2244	TGGGCAATCGCTAACAGAGATTCCAGACGGCTCCTTCTCGACGACAGAGTGGACACTACCG	2303
QY	2571	GCCGTTCCAGGGCACATGGGGGGTCTGTCTCTCCCTCCGACCCACCATATGCTCCAGGGC	2630
Db	2304	GCCGTTCCAGGGCACATGGGGGGTCTGTCTCTCCCTCCGACCCACCATATGCTCCAGGGC	2363
QY	2631	CCCAACTGCAAAAGTTGGCCCACTGAGAGACCACTCCTGCGTCCCTCTGAGTCTGAAAG	2690
Db	2364	CCCAACTGCAAAAGTTGGCCCACTGAGAGACCACTCCTGCGTCCCTCTGAGTCTGAAAG	2423
QY	2691	TGAACCGTACACCTTTCGCCATCCCAACAAATGGGGATGTAAAGACAGAAAGACACAGACAT	2750
Db	2424	TGAACCGTACACCTTTCGCCATCCCAACAAATGGGGATGTAAAGACAGAAAGACACAGACAT	2483
QY	2751	TCCCTTACTGAAACACTCAGAGACGCCATGAGAGCCAGACAGATAACTGTATCCATTCCAGAC	2810
Db	2484	TCCCTTACTGAAACACTCAGAGAGGCCATGAGAGCCAGACAGATAACTGTATCCATTCCAGAC	2543
QY	2811	GCTTGGCTGAGTTTCATTAAGACAGGGGACCTGAGACACCCGTCGGTGTCTTAACCAAGAA	2870
Db	2544	GCTTGGCTGAGTTTCATTAAGACAGGGGACCTGAGACACCCGTCGGTGTCTTAACCAAGAA	2603
QY	2871	TCCTAAGAAGAGGAATTTATACAAAGGAAGACAGAGAGGTTTTCTGTGACACCGCCAAC	2930
Db	2604	TCCTAAGAAGAGGAATTTATACAAAGGAAGACAGAGAGGTTTTCTGTGACACCGCCAAC	2663
QY	2931	TTACATTTGCTCACTGGACACTCATTTTAAGGGCAAGACATTGAAAAATGATGAATTCACATC	2990
Db	2664	TTACATTTGCTCACTGGACACTCATTTTAAGGGCAAGACATTGAAAAATGATGAATTCACATC	2723
QY	2991	TGGATACAGTATATACAGCTCATGTGTCTCTCAACTTAAGGCTGTGCGGTAGCCACAGCTG	3050
Db	2724	TGGATACAGTATATACAGCTCATGTGTCTCTCAACTTAAGGCTGTGCGGTAGCCACAGCTG	2783
QY	3051	TAAAGAGAGGAGAGAGCCGTGAGTACACCTGTAGAGGTTGGACGAAGCCCTGGATTACAG	3110
Db	2784	TAAAGAGAGGAGAGAGCCGTGAGTACACCTGTAGAGGTTGGACGAAGCCCTGGATTACAG	2843
QY	3111	AGTGTAAACAGAGGCTTGGCCTCTTCAGGACAAAGTTTCCAAATTCGAAGGAGCCTACCT	3170
Db	2844	AGTGTAAACAGAGGCTTGGCCTCTTCAGGACAAAGTTTCCAAATTCGAAGGAGCCTACCT	2903
QY	3171	GAGGTCCCTACTCTCACTGGGGGTCCGAGGATTAAGACACAATGTGCCCTTTTATTATTT	3230
Db	2904	GAGGTCCCTACTCTCACTGGGGGTCCGAGGATTAAGACACAATGTGCCCTTTTATTATTT	2963
QY	3231	ATTATTTTGGGGGCGGTGTATTTTAAGAGATCAAAATGTATTAACACACTTACGCTTTTC	3290
Db	2964	ATTATTTTGGGGGCGGTGTATTTTAAGAGATCAAAATGTATTAACACACTTACGCTTTTC	3023
QY	3291	ACCTGACTAGTATATACATCATCTAACTGATGGTTGGATCCCTGGGTTGTGACTTCTACTG	3350

```
|||||
Db 3024 ACCGACTGTGATTAACATACTACTAGTGGATGCTGGGTGTGCTTCTACTG 3083
OY 3351 ACCGCTGATTAACGTGTGCTGTCCCAAGGTGGGTAATTAATTAACATCTGTCCA 3410
Db 3084 ACCGCTGATTAACGTGTGCTGTCCCAAGGTGGGTAATTAATTAACATCTGTCCA 3413
OY 3411 CCAGAAAAAGATGTGTGTTTGAGAGCATTGACATATCTGCTTGTGATAGAGACT 3470
Db 3144 CCAGAAAAAGATGTGTGTTTGAGAGCATTGACATATCTGCTTGTGATAGAGACT 3203
OY 3471 CCGATTCCTAGGTGCGTTCGTGTTATGCCATTTGTGGAATTCATCTTGAATCCATT 3550
Db 3204 CCGATTCCTAGGTGCGTTCGTGTTATGCCATTTGTGGAATTCATCTTGAATCCATT 3263
OY 3531 GTCCTATAGTCTAGCATTAAGAGAAATTTCTCAATTTCCATGTGCGGTTCTCTAGC 3590
Db 3264 GTCCTATAGTCTAGCATTAAGAGAAATTTCTCAATTTCCATGTGCGGTTCTCTAGC 3323
OY 3591 TGCAGCAATCTTTGACATTTAAAGAGAAATTTAGAGAAATTTCTCATCTCTAAAAATG 3650
Db 3324 TGCAGCAATCTTTGACATTTAAAGAGAAATTTAGAGAAATTTCTCATCTCTAAAAATG 3383
OY 3651 TTTTAATATATACCAACAGTGGGCCCTGCTATGTTTCTGTGGCAGCTGCAACCCAT 3710
Db 3384 TTTTAATATATACCAACAGTGGGCCCTGCTATGTTTCTGTGGCAGCTGCAACCCAT 3443
OY 3711 TACTTGTAGTCTTAAACAAACACATTAAGCTTATAGTCTGCGGATGAGAAATTCAGAAAT 3770
Db 3444 TACTTGTAGTCTTAAACAAACACATTAAGCTTATAGTCTGCGGATGAGAAATTCAGAAAT 3503
OY 3771 GGATGTCCCTGATGAATAATCAGAGTGTGACAGAGCTGTGCTCTCTGAGAGCTCTAG 3830
Db 3504 GGATGTCCCTGATGAATAATCAGAGTGTGACAGAGCTGTGCTCTCTGAGAGCTCTAG 3563
OY 3831 GGAGAAGCCGTTCTGCTGCAATTTCAAGCTTCTAGAGCTGCGTGTGCTTCCAGAGCTCCA 3890
Db 3564 GGAGAAGCCGTTCTGCTGCAATTTCAAGCTTCTAGAGCTGCGTGTGCTTCCAGAGCTCCA 3623
OY 3891 GTGCGTGTGCAAGCTTTCTCAGATGAGCATGAGACTGTGCACTGCGCTCCACTTCCCTC 3950
Db 3624 GTGCGTGTGCAAGCTTTCTCAGATGAGCATGAGACTGTGCACTGCGCTCCACTTCCCTC 3663
OY 3951 TTTGACTTTACAAAGCCACAGAGAGATCCAGATTAATCTCTCCATCTAAAGATCTCTCA 4010
Db 3684 TTTGACTTTACAAAGCCACAGAGAGATCCAGATTAATCTCTCCATCTAAAGATCTCTCA 3743
OY 4011 TCATCTCGGAAGAGCCCTTTTGCCATTCGAGAGACATAGCCACAGGTGGGATTAAGACC 4070
Db 3744 TCATCTCGGAAGAGCCCTTTTGCCATTCGAGAGACATAGCCACAGGTGGGATTAAGACC 3803
OY 4071 AGGACATCTTTGGGGTCTGTATTCTGCTTACACACCTTCCGACACBAGACTCCACA 4130
Db 3804 AGGACATCTTTGGGGTCTGTATTCTGCTTACACACCTTCCGACACBAGACTCCACA 3863
OY 4131 GGAGAGGCTCAAAATATGATGTGGGCAAGAGATGTTTGTATTAGCTTGGGACTCTAAC 4190
Db 3864 GGAGAGGCTCAAAATATGATGTGGGCAAGAGATGTTTGTATTAGCTTGGGACTCTAAC 3923
OY 4191 ACTTAAAAAACCACATCAGAAATCTGGCCATGCTGGGGCTCAACATTTCTACCTAGC 4250
Db 3924 ACTTAAAAAACCACATCAGAAATCTGGCCATGCTGGGGCTCAACATTTCTACCTAGC 3983
OY 4251 AACAACTGGGTGAGTGGGACCAAGCTGTGCTTTAGAAAGGGGTGTCACCTTACCAGG 4310
Db 3984 AACAACTGGGTGAGTGGGACCAAGCTGTGCTTTAGAAAGGGGTGTCACCTTACCAGG 4043
OY 4311 TCACCAACAGCCACATAGGCGCTATCATCTCCACATAGAGGCTAAGTGTGTTCTTA 4370
Db 4044 TCACCAACAGCCACATAGGCGCTATCATCTCCACATAGAGGCTAAGTGTGTTCTTA 4103
OY 4371 CTGATCAATGCCCTGAGGTGTGATTTATGAAAAAGAAACAGTGGATTAATC 4430
|||||

|||||
Db 4104 CTGATCAATGCCCTGAGGTGTGATTTATGTAATGAAAAAGAAAGAGTGAATATC 4163
OY 4431 TCTATCAGGTGATAGACCATGAGACCAATGTGTGCTACATTAACCTTTTCTTTT 4490
Db 4164 TCTATCAGGTGATAGACCATGAGACCAATGTGTGCTACATTAACCTTTTCTTTT 4223
OY 4491 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4550
Db 4224 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4283
OY 4551 TGGAGTCAAGTGGGCAATCTCGGCTCAGTGCACATCTGCTCTGCGGCTCAAGCAAT 4610
Db 4284 TGGAGTCAAGTGGGCAATCTCGGCTCAGTGCACATCTGCTCTGCGGCTCAAGCAAT 4353
OY 4611 CTCACACCTCAGCCTCCCAATATAGCTGGATCATGTGACACAAACACCATGCCCCAGTAA 4670
Db 4344 CTCACACCTCAGCCTCCCAATATAGCTGGATCATGTGACACAAACACCATGCCCCAGTAA 4403
OY 4671 TTTTGTATTTTGTGTAGAGACAGGTTTACCATGTTGCCAGGCTGCTCAACCTCT 4730
Db 4404 TTTTGTATTTTGTGTAGAGACAGGTTTACCATGTTGCCAGGCTGCTCAACCTCT 4463
OY 4731 GGGCTCAAGCAATCTGCTGCTGGGCTCCCAAGTGGTGGGATTAAGATGAGGCA 4790
Db 4464 GGGCTCAAGCAATCTGCTGCTGGGCTCCCAAGTGGTGGGATTAAGATGAGGCA 4523
OY 4791 CCGCATCCAGCCCCACACCCCTCATTTATACCAATTAACCTGCGCAGTAATGTGAGCTTT 4850
Db 4524 CCGCATCCAGCCCCACACCCCTCATTTATACCAATTAACCTGCGCAGTAATGTGAGCTTT 4583
OY 4851 GCTTCTCAGCCCTGCTGTGATCTGGAAGAGAGGATTAATGTAAGCTTGTACAGCA 4910
Db 4584 GCTTCTCAGCCCTGCTGTGATCTGGAAGAGAGGATTAATGTAAGCTTGTACAGCA 4643
OY 4911 GTCCCAAGTCAATATTTGTGCGGCAAAAACCTTCTCAAAAATTAATGACTCTAATG 4970
Db 4644 GTCCCAAGTCAATATTTGTGCGGCAAAAACCTTCTCAAAAATTAATGACTCTAATG 4703
OY 4971 TATTCATATGAATTCACCTTGAATAATGACAGCCCTCAACTTGTTCACATAGCATTAATGAA 5030
Db 4704 TATTCATATGAATTCACCTTGAATAATGACAGCCCTCAACTTGTTCACATAGCATTAATGAA 4763
OY 5031 AGGAATTTTATAGTCTCTTAATAGGCTGTATCTGCAAGACCTCTTGAACACTTTCAGAG 5090
Db 4764 AGGAATTTTATAGTCTCTTAATAGGCTGTATCTGCAAGACCTCTTGAACACTTTCAGAG 4823
OY 5091 GATGGAATTTAATGATGATGCTTGGGCTTGGCTATGAGCACTTCCCTCTGAAAGTC 5150
Db 4824 GATGGAATTTAATGATGATGCTTGGGCTTGGCTATGAGCACTTCCCTCTGAAAGTC 4883
OY 5151 TGGTCTGCGCCAGTGAACCTTGGCTTGTGAGCGAGATGCTGACCTGCATTAAGGGC 5210
Db 4884 TGGTCTGCGCCAGTGAACCTTGGCTTGTGAGCGAGATGCTGACCTGCATTAAGGGC 4943
OY 5211 CAAAGAGGGCTGCGGCTTCCCTTCCACTGAAGAGCCCTTATTTGAATTCACCTGTGTG 5270
Db 4944 CAAAGAGGGCTGCGGCTTCCCTTCCACTGAAGAGAGCCCTTATTTGAATTCACCTGTGTG 5003
OY 5271 GAGGCTTAAGCCCTGCAATTCGAGATGCCAACCTGCCAGGCCCTTCAAGAGAGGACTA 5330
Db 5004 GAGGCTTAAGCCCTGCAATTCGAGATGCCAACCTGCCAGGCCCTTCAAGAGAGGACTA 5063
OY 5331 GGTGCTGCTGATTCACCCCAAGGTGGGATTTGGCTTCTTAAGGTGGCTACTTGTACCA 5390
Db 5064 GGTGCTGCTGATTCACCCCAAGGTGGGATTTGGCTTCTTAAGGTGGCTACTTGTACCA 5123
OY 5391 TCACCAACATCACTGTTGCTGCAAGAGACACACAGTGGCAATTTCTTCAACTGAGGGC 5450
Db 5124 TCACCAACATCACTGTTGCTGCAAGAGACACACAGTGGCAATTTCTTCAACTGAGGGC 5183
OY 5451 TCAAAATCTCTGAGCAAGTGTGCTGCTGAGACAGATTAATTCCTGGAGCTGTGCTCA 5510
Db 5184 TCAAAATCTCTGAGCAAGTGTGCTGCTGAGACAGATTAATTCCTGGAGTGTGCTCA 5243
|||||
```

5511 GTGAAGGGGGCCAGCCTGAGAACCTGCTCTTTCTTTAAAGCCAGGCCACTTAC 5570
|||||
5244 GTGAAGGGGGCCAGCCTGAGAACCTGCTCTTTCTTTAAAGCCAGGCCACTTAC 5503
|||||
5571 ATAAACATTTTCAGGCTCAGTGAAGAGTGAAGTGCATTTGTGAAGCTTACTGATG 5630
|||||
5304 ATAAACATTTTCAGGCTCAGTGAAGAGTGAAGTGCATTTGTGAAGCTTACTGATG 5563
|||||
5631 CCAGCCCACTGCTCATTCACAGTGTGCTGCTGCTACAGAGAAAGGCCAGCGCATGAG 5690
|||||
5364 CCAGCCCACTGCTCATTCACAGTGTGCTGCTGCTACAGAGAAAGGCCAGCGCATGAG 5423
|||||
5691 ACTGCTCTTAATGCTGTGTCTATTCACAGAGAAAGGAGTGTCAAGAGAGTCAACT 5750
|||||
5424 ATGGCTCTTAATGCTGTGTCTATTCACAGAGAAAGGAGTGTCAAGAGAGTCAACT 5483
|||||
5751 GGGACACAGCAACAGCCAGCCAGCATGCTGTGTAAGGTTAGCAGACTGTGTGTG 5810
|||||
5484 GGGACACAGCAACAGCCAGCCAGCATGCTGTGTAAGGTTAGCAGACTGTGTGTG 5543
|||||
5811 GATCTCAGTGTCTCAGTGAAGTAAATTTATTCATTGACAGATACCTTTTGTGCAATT 5870
|||||
5544 GATCTCAGTGTCTCAGTGAAGTAAATTTATTCATTGACAGATACCTTTTGTGCAATT 5803
|||||
5871 TATTCATTCTGCTGCTTTAAATTAACAAATGTAACAAACAAACAAACAACTATCACTGTTTA 5930
|||||
5604 TATTCATTCTGCTGCTTTAAATTAACAAATGTAACAAACAAACAAACAACTATCACTGTTTA 5663
|||||
5931 AGTCTCTCGGCTACTGTGCTCCCTGCTTTCAGTAGAGCCCGCGTGTCCAGTGTGTACTG 5990
|||||
5664 AGTCTCTCGGCTACTGTGCTCCCTGCTTTCAGTAGAGCCCGCGTGTGTCCAGTGTGTACTG 5723
|||||
5991 TGACAGGCTCAGCATGGGCTCAGCAGATGCTGCTTAAATTTGTGATGATACAGAAAGCC 6050
|||||
5724 TGACAGGCTCAGCATGGGCTCAGCAGATGCTGCTTAAATTTGTGATGATACAGAAAGCC 5783
|||||
6051 AGGCTTTGGGATTCAGAGTCTTCTCTTCATTTGATGATGATGATGATGATGATGATGATGAT 6110
|||||
5784 AGGCTTTGGGATTCAGAGTCTTCTCTTCATTTGATGATGATGATGATGATGATGATGATGAT 5843
|||||
6111 GTTTTGTCCGGAATTAATTAATAGTCTTGGAGTCTGCGCAAAAAA 6163
|||||
3844 GTTTTGTCCGGAATTAATTAATAGTCTTGGAGTCTGCGCAAAAAA 5896
|||||

RESULT 4
AK023834 5573 bp mRNA linear PRI 01-AUG-2002
LOCUS AK023834 Homo sapiens cDNA FLJ13772 fls, clone PLACE4000300.
DEFINITION AK023834
ACCESSION AK023834.1 GI:10435889
VERSION AK023834.1
KEYWORDS Oligo capping; fls (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone lib:PLACE4
clone:PLACE4000300.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
Arima,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Negahari,K., Masuko,Y. and Oshima,A.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5573)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2000) Takeo Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source

location/Qualifiers

1..5573

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="PLACE4000300"

/issue_type="placenta"

/clone_lib="PLACE4"

/note="cloning vector: pMR185FL3"

275..2224

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14695.1"

/db_xref="GI:10435890"

/translation="MOEGYKMLHLPHFPRNVSFSTANRSSIKRLCTIESFEGE
SATLSANTPEGPDELMTCQFVFAHLRASVFLNLSNCKEREVEYITPEST
TNEVEFKLEDKQDPGNMAGNFNLSQCDQDADSPGLRLQFVLVQHPONESKITYV
DLSNERAMSLITIEPRVKQSRKFPQCFVCLSRCSNLTLSGSKHISFLCDLT
RLMNVEXTISCDBHRYCQKRSYLOVPDILHLPVELHDFSKILVPRDLSILVLP
AOKLOOHEKPCNTSFSYLVASAIPDSOLLYFSRCPGSGITQIOYKONISYLTREA
PSFROASROGLTVSFTIPIYFKREGVTPYTPDKSKYILTPPMNDRGLPBLTSVSNMIS
VPDQVACILFFERKSGVCOQTRAFMIIQEORTRAEELFSLDEVLPRFHHHSFW
VINSQSPISGKQDLDFSVTLTPRTVDLTVLLAVGGVLLLSLGLITCVKRRK
KLTNKGPAVGIYNGNINTEMPRPKFKQGRKDNDSHVAVIEDTAVYGHLLQDSGS
FLOPEVDYTRPFGQVPCPSPPGKPTCSAPAKLATEBPSPSESESEPTFSHP
NNGDVSSKRDILPILOTQEMPEPAE"

BASE COUNT 1373 a 1524 c 1290 g 1386 t

ORIGIN

Query Match 89.5%; Score 5518.4; DB 9; Length 5573;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 5566; Conservative 0; Mismatches 6; Indels 4; Gaps 4;629 CGTTGTGGCTACCCCTCAAGAGAACTTTCATCTGGGATGTCAAAGCTCATAGAGCATG 688
|||||
61 CGTTGTGGCTACCCCTCAAGAGAACTTTCATCTGGGATGTCAAAGCTCATAGAGCATG 120
|||||
689 GTTAGAGCTGAGTTTCCATCCCTCGCTGAGGAGATCGGTCGGGTGAGAGCTGCC 748
|||||
121 G-TTAGAGCTGAGTTTCCATCCCTCGCTGAGGAGATCGGTCGGGTGAGAGCTGCC 179
|||||
749 CAGACGAGTCACTCACTCATCAGCGCGCGAATCATGTCACCCGTTGATGATGCGGA 808
|||||
180 CAGACGAGTCACTCACTCATCAGCGCGCGAATCATGTCACCCGTTGATGATGCGGA 239
|||||
809 CATTGACAGCAATGGCAGTGTGCCCGGATCAAGATGCAAGAGAGTGAATAGGCT 868
|||||
240 CATTGACAGCAATGGCAGTGTGCCCGGATCAAGATGCAAGAGAGTGAATAGGCT 299
|||||
869 TACACCTCCCATGGTTCACCCAGAAATGTCGCCGCTTCAGCATTTGCAAAACCGTCAT 928
|||||
300 TACACCTCCCATGGTTCACCCAGAAATGTCGCCGCTTCAGCATTTGCAAAACCGTCAT 359
|||||
929 CTTTAAACCTGTGTGATCATGAGTGTGTGTTGAGGTTGAAGGCTCAGCAACCTGCA 988
|||||
360 CTTTAAACCTGTGTGATCATGAGTGTGTGTTGAGGTTGAAGGCTCAGCAACCTGCA 419
|||||
989 TGTCTGCAACTACCCAGAAAGGCTTCCTGAGAGTGAAGTCAAGAGCTGAGCTGTTGTCG 1048
|||||
420 TGTCTGCAACTACCCAGAAAGGCTTCCTGAGAGTGAAGTCAAGAGCTGAGCTGTTGTCG 479
|||||
1049 TTCTCTGACACCTGCGGGCCAGGCTCTCTTCTCAACTTCAACCTTCCAACTGTGAGA 1108
|||||
480 TTCTCTGACACCTGCGGGCCAGGCTCTCTTCTCAACTTCAACCTTCCAACTGTGAGA 539
|||||

QY 1109 GGAAGGAGGAGCGGGTTGAATCTACATCCGGGCTCCACCACCAACCCGAGGTGTTC 1188
|||||
Db 540 GGAAGGAGGAGCGGGTTGAATCTACATCCGGGCTCCACCACCAACCCGAGGTGTTC 599
QY 1169 AGCTGGAGGACACGAGCTTGGGAACATGGCGGGGAACCTTCAACCTCTCTGCAAGGCT 1228
|||||
Db 600 AGCTGGAGGACACGAGCTTGGGAACATGGCGGGGAACCTTCAACCTCTCTGCAAGGCT 659
QY 1229 GTGACCAAGATGCCCAAGTCCAGGATCCCTCCGGCTGCACTTCCAAATTTTGGTCCAA 1288
|||||
Db 660 GTGACCAAGATGCCCAAGTCCAGGATCCCTCCGGCTGCACTTCCAAATTTTGGTCCAA 719
QY 1289 ATCCACAATAATGAAAGCAATAAATCTACGTGTTGACTTGAGTAATGAGCGAGCCATGT 1348
|||||
Db 720 ATCCACAATAATGAAAGCAATAAATCTACGTGTTGACTTGAGTAATGAGCGAGCCATGT 779
QY 1349 CACACACATGAGCAGCAGCGGCGCTCAAAACAGACCCCAAGTTGTCCTCGGCTTTCCG 1408
|||||
Db 780 CACTCACCATGAGCCACGGCCGCTCAAAACAGACCCCAAGTTGTCCTCGGCTTTCCG 839
QY 1409 TGTGTCTAGATCTCGGACCTGAGTACCAACCTCACCTGACATCTGGCTCCAAACACA 1468
|||||
Db 840 TGTGTCTAGATCTCGGACCTGAGTACCAACCTCACCTGACATCTGGCTCCAAACACA 899
QY 1469 AAATCTCCTTCTTTGTGATGATCTGAACGCTCTGTGAGTGAATGTGAAAAAACATTA 1528
|||||
Db 900 AAATCTCCTTCTTTGTGATGATCTGAACGCTCTGTGAGTGAATGTGAAAAAACATTA 959
QY 1529 GCTGCACAGACCAACGGGTAAGTCCAAAGAAATCTACTCATCTCAGTGGCCGCTACACA 1588
|||||
Db 560 GCTGCACAGACCAACGGGTAAGTCCAAAGAAATCTACTCATCTCAGTGGCCGCTACACA 1019
QY 1589 TCTCTCACTGCTGCTGTGAGAGTGCATGACTTCTCTGGAAGCTCTGTGCTCCAAAGACA 1648
|||||
Db 1020 TCTCTCACTGCTGCTGTGAGAGTGCATGACTTCTCTGGAAGCTCTGTGCTCCAAAGACA 1079
QY 1649 GGGTCACGCTGTGCTGTGCTGCAAGCCCAAGAAAGCTGCAAGCAATACACAGAGAACCTT 1708
|||||
Db 1080 GGGTCACGCTGTGCTGTGCTGCAAGCCCAAGAAAGCTGCAAGCAATACACAGAGAACCTT 1139
QY 1709 GCAACACACAGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1768
|||||
Db 1140 GCAACACACAGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
QY 1769 GCTCCTTCTGCGCGGAGGCTCTATCAAGCAGATCCAGGTGAACAGAACTCTCGGTGA 1828
|||||
Db 1200 GCTCCTTCTGCGCGGAGGCTCTATCAAGCAGATCCAGGTGAACAGAACTCTCGGTGA 1259
QY 1829 CCTTCCGACCTTTGGCCCCAGCTTCCAAACAAAGGCTTCCAGGAGGCTTGAACGCTGT 1888
|||||
Db 1260 CCTTCCGACCTTTGGCCCCAGCTTCCAAACAAAGGCTTCCAGGAGGCTTGAACGCTGT 1319
QY 1889 CCTTATACCTTATTTCAAAAGAGAGGCGTTTTCACGCTGACCCCTGACACAAAGACA 1948
|||||
Db 1320 CCTTATACCTTATTTCAAAAGAGAGGCGTTTTCACGCTGACCCCTGACACAAAGACA 1379
QY 1949 AGCTTACCTGAGAGACCCCAACTGGAGACCGGGGCTGCCATCCCTCCTGCTGTCT 2008
|||||
Db 1380 AGCTTACCTGAGAGACCCCAACTGGAGACCGGGGCTGCCATCCCTCCTGCTGTCT 1439
QY 2009 GGAACATACAGCTGCTCCAGAGACAGTGGCTGCTGACTTTTATTAAGAGCGGAGCG 2068
|||||
Db 1440 GGAACATACAGCTGCTCCAGAGACAGTGGCTGCTGACTTTTATTAAGAGCGGAGCG 1499
QY 2069 GCGTGTGCTGCAAGAGAGGCGGCTTCAATGATCATCCAGAGACAGCGGAGCGGCTG 2128
|||||
Db 1500 GCGTGTGCTGCAAGAGAGGCGGCTTCAATGATCATCCAGAGACAGCGGAGCGGCTG 1559
QY 2129 AGGAGATCTTACGCTGAGAGAGAGTGTCTCCCAAGCCAAAGCTTCCACATCACAGCT 2188
|||||
Db 1560 AGGAGATCTTACGCTGAGAGAGAGTGTCTCCCAAGCCAAAGCTTCCACATCACAGCT 1619

QY 2189 TCTGGGTCAACATCTCTAATCTGACGCCCCACAGAGCGGCAACACAGTACCTGCTCTTCT 2248
|||||
Db 1620 TCTGGGTCAACATCTCTAATCTGACGCCCCACAGAGCGGCAACACAGTACCTGCTCTTCT 1679
QY 2249 CGGTGACACTTACCCCAAGGACTGTGACCTGACCTGATCCTCATGCTCAGCGGGGAG 2308
|||||
Db 1680 CGGTGACACTTACCCCAAGGACTGTGACCTGACCTGATCCTCATGCTCAGCGGGGAG 1739
QY 2309 GTGAGTCTTACTGCTGTGCTGCTGCGGCTCATCATTTGCTGTGAAAAAGAAAA 2368
|||||
Db 1740 GTGAGTCTTACTGCTGTGCTGCTGCGGCTCATCATTTGCTGTGAAAAAGAAAA 1799
QY 2369 AGAAGACAACAAAGGCGCCGCTGTGGTATCTCAATGGCAATCATCTAGATGC 2428
|||||
Db 1800 AGAAGACAACAAAGGCGCCGCTGTGGTATCTCAATGGCAATCATCTAGATGC 1859
QY 2429 CGAGGAGCCCAAAAAGTTTGAAGAAAGGCGCAAGGCAATGACCTCCATGTATGAG 2488
|||||
Db 1860 CGAGGAGCCCAAAAAGTTTGAAGAAAGGCGCAAGGCAATGACCTCCATGTATGAG 1919
QY 2489 TCATCGAGGACACCATGGTATATGGGATCTGCTACAGAGATTCCAGCGGCTCTTCCCTG 2548
|||||
Db 1920 TCATCGAGGACACCATGGTATATGGGATCTGCTACAGAGATTCCAGCGGCTCTTCCCTG 1979
QY 2549 AGCCAGAGGTGACACCTACCGGCGCTTCCAGGCGACCATGGGGGTCTGTCTCCCTCC 2608
|||||
Db 1980 AGCCAGAGGTGACACCTACCGGCGCTTCCAGGCGACCATGGGGGTCTGTCTCCCTCC 2039
QY 2609 CACCCACATATGCTCCAGGGCGCCCACTGCAAGTTGGGCCACTGAGAGGACACCTCTC 2668
|||||
Db 2040 CACCCACATATGCTCCAGGGCGCCCACTGCAAGTTGGGCCACTGAGAGGACACCTCTC 2099
QY 2669 GCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCCCAATGGGATG 2728
|||||
Db 2100 GCTCCCTCTCTGAGTCTGAGAGTGAACCGTACACCTTCTCCATCCCAATGGGATG 2159
QY 2729 TAACGACAGGACACAGACTTCTCTTACTGAACACTCAGAGACCCATGAGCCACGAG 2788
|||||
Db 2160 TAACGACAGGACACAGACTTCTCTTACTGAACACTCAGAGACCCATGAGCCACGAG 2219
QY 2789 AATATCTTGAATCCATTCACAGAGCTTGTGCTGAGTTTCTAAAGCAGGCGCATGAGAC 2848
|||||
Db 2220 AATATCTTGAATCCATTCACAGAGCTTGTGCTGAGTTTCTAAAGCAGGCGCATGAGAC 2279
QY 2849 GGTCCGTCTCTTAACAGAAATCTTAAAGAGGAAATTAACAGAGAAACAGCAG 2908
|||||
Db 2280 GGTCCGTCTCTTAACAGAAATCTTAAAGAGGAAATTAACAGAGAAACAGCAG 2339
QY 2909 GGTTCCTGACACCGCCCAACTTCAATGCTCAATGCTCAATGCTCAATGCTCAATG 2968
|||||
Db 2340 GGTTCCTGACACCGCCCAACTTCAATGCTCAATGCTCAATGCTCAATGCTCAATG 2399
QY 2969 TTGAATAATGATGAATTCATCTGATACAGTCAATGCTCAATGCTCAATGCTCAAT 3028
|||||
Db 2400 TTGAATAATGATGAATTCATCTGATACAGTCAATGCTCAATGCTCAATGCTCAAT 2459
QY 3029 GGTGTGCGGTTAGCAGCAGCTGTATAGAGAGAGAGGCGCTGAGTCACTAGCATAGG 3088
|||||
Db 2460 GGTGTGCGGTTAGCAGCAGCTGTATAGAGAGAGAGGCGCGAGTCACTAGCATAGG 2519
QY 3089 TTGACAGCAAGCCCTGTGATTCAGAGTGTAAACAGAGGCTTCCCTTTCAGAGAACAGT 3148
|||||
Db 2520 TTGACAGCAAGCCCTGTGATTCAGAGTGTAAACAGAGGCTTCCCTTTCAGAGAACAGT 2579
QY 3149 TTCATATTCAGAGGAGCTTACCTGAGGCTTACCTTACTTGGGCTCCCAAGATGAAG 3208
|||||
Db 2580 TTCATATTCAGAGGAGCTTACCTGAGGCTTACCTTACTTGGGCTCCCAAGATGAAG 2639
QY 3209 GACAATGTGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3268
|||||
Db 2640 GACAATGTGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2699
QY 3269 GTATTAACCACTTACCTTTTTCACCTGACTTATTAATTAATTAATTAATTAATTAAT 3328


```
|||||
Db 2700 GTATACCACTAGCTCTTTTACGACTACTATTACTCTACTACTGTTGGAT 2759
QY 3329 GCGTGGGTGTGACTTCTACTGACCGCTAGATAAACGTGTGCTGTCCCGAGTGTG 3388
Db 2760 GCTGTGGTGTGACTTCTACTGACCGCTAGATAAACGTGTGCTGTCCCGAGTGTG 2819
QY 3389 GAAATATTACATCTGTCCAAACGAAAGATGTGTGTGTGAGCAGCATTTGACACA 3448
Db 2820 GAATTAATTACATCTGTCCAAACGAAAGATGTGTGTGTGAGCAGCATTTGACACA 2879
QY 3449 TATCTCTTGTGATAGAGACTTCTGATTCCTAGTCTGAGTGGTGTGATTTCCATTGTG 3508
Db 2880 TATCTCTTGTGATAGAGACTTCTGATTCCTAGTCTGAGTGGTGTGATTTCCATTGTG 2939
QY 3509 GAAATTCATCTTGAATTCCTCATTTGTCTATAGTCTAGCAATAAGAGAAATTTCTCAAGT 3568
Db 2940 GAAATTCATCTTGAATTCCTCATTTGTCTATAGTCTAGCAATAAGAGAAATTTCTCAAGT 2999
QY 3569 TTCCATGTGGGGTTCCTAGCTGACGCAATACTTGACATTTAAGAGAAATTTAGAGA 3628
Db 3000 TTCCATGTGGGGTTCCTAGCTGACGCAATACTTGACATTTAAGAGAAATTTAGAGA 3059
QY 3629 ATATTCTCATCTCTAAAAATGTTTAAATATATACCAACAGTGGCCCTGCATTAGTT 3688
Db 3060 ATATTCTCATCTCTAAAAATGTTTAAATATATACCAACAGTGGCCCTGCATTAGTT 3119
QY 3689 TTCTGTTCGCACTGCAACCCATTAATGTTGGTAGCTTAAAAACAACATTAAGTTATAGTC 3748
Db 3120 TTCTGTTCGCACTGCAACCCATTAATGTTGGTAGCTTAAAAACAACATTAAGTTATAGTC 3179
QY 3749 CTGGGGATCAGAAATTCAAAATGGATGTCCTCCTAAATGAAATTCAGGTGCACACAGCT 3808
Db 3180 CTGGGGATCAGAAATTCAAAATGGATGTCCTCCTAAATGAAATTCAGGTGCACACAGCT 3239
QY 3809 GTGCTCTCTTCTGAGAGCTCTAGAGAGAACGCGTCTCTGCGCATTTCAAGCTTCTAGAG 3868
Db 3240 GTGCTCTCTTCTGAGAGCTCTAGAGAGAACGCGTCTCTGCGCATTTCAAGCTTCTAGAG 3299
QY 3869 CTGCTGTCATCTCCAGAGCTCCAGTGCCTGTGTCAGACTTTTCTCATGTGCATCTGTGA 3928
Db 3300 CTGCTGTCATCTCCAGAGCTCCAGTGCCTGTGTCAGACTTTTCTCATGTGCATCTGTGA 3359
QY 3929 CACTGGCCCTCCACTTCCCTCTTTGACTTACAAAGCCCAAGGAAAGATCCAGGATTAAT 3988
Db 3360 CACTGGCCCTCCACTTCCCTCTTTGACTTACAAAGCCCAAGGAAAGATCCAGGATTAAT 3419
QY 3989 CTCTCCATCTAAAGATCTTTCATCATCTGGAAGAGCTTTTGCCATGCAAGCAACATA 4048
Db 3420 CTCTCCATCTAAAGATCTTTCATCATCTGGAAGAGCTTTTGCCATGCAAGCAACATA 3479
QY 4049 GCCACAGGTGGGATTAGACACAGACATCTTTGGGGTGTCTGTTATTTCTGCTACACAC 4108
Db 3480 GCCACAGGTGGGATTAGACACAGACATCTTTGGGGTGTCTGTTATTTCTGCTACACAC 3539
QY 4109 CTTCCTGCACTACTTCCACAGAGAGGCTACAAAATGATCTGGGCGACAGGAGATGTTT 4168
Db 3540 CTTCCTGCACTACTTCCACAGAGAGGCTACAAAATGATCTGGGCGACAGGAGATGTTT 3599
QY 4169 TGTTTAGCTTTGGGAGCTTAACACTT - AAAAAAAAACCCGAATCAGAAAGATCTGGGCAATG 4227
Db 3600 TGTTTAGCTTTGGGAGCTTAACACTTAAAAAAAACCCGAATCAGAAAGATCTGGGCAATG 3659
QY 4228 TGGGGCTCACAATCTCTACCTAGCAACAATGCTGTGAGCTGGGCGACAGCTGTGCCTTTA 4287
Db 3660 TGGGGCTCACAATCTCTACCTAGCAACAATGCTGTGAGCTGGGCGACAGCTGTGCCTTTA 3719
QY 4288 GAAAGGGGTGTCACTTACAGAGGTCAACACAGCCACACTAGCCCTATCATTCCACACA 4347
Db 3720 GAAAGGGGTGTCACTTACAGAGGTCAACACAGCCACACTAGCCCTATCATTCCACACA 3779
QY 4348 ATGAGGCTAAGTGTGTTGTTTCTACTATCAATGCCCTGTGAGTGTGATTTATTTGTAAG 4407
|||||

Db 3780 ATGAGGCTAAGTGTGTTGTTTCTACTATCAATGCCCTGTGAGTGTGATTTATTTGTAAG 3839
QY 4408 AAAAAAGAGCTGGGATTAATCTTAATAGGTGAGTGAACCATGAGCAATATGTGTG 4467
Db 3840 AAAAAAGAGCTGGGATTAATCTTAATAGGTGAGTGAACCATGAGCAATATGTGTG 3899
QY 4468 TCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4527
Db 3900 TCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3959
QY 4528 GGATCTCATCTCTGTGCTAGGCTGAGGTGACAGTGGCGCAATCTCGGCTCAGCAACT 4587
Db 3960 GGATCTCATCTCTGTGCTAGGCTGAGGTGACAGTGGCGCAATCTCGGCTCAGCAACT 4019
QY 4588 CTGCTCTCTGGGCTCAAGCAATTTCTCCACCTCAGCCTCCCAATATAGCTGGGATCACTG 4647
Db 4020 CTGCTCTCTGGGCTCAAGCAATTTCTCCACCTCAGCCTCCCAATATAGCTGGGATCACTG 4079
QY 4648 CACAAACCACTAGCCAGCTAATTTTGTATTTTGTAGAGACAGGGTTTCAACATGTT 4707
Db 4080 CACAAACCACTAGCCAGCTAATTTTGTATTTTGTAGAGACAGGGTTTCAACATGTT 4138
QY 4708 GCCCAGGCTGTCTCAACCTCTGTGGCTCAAGCAATCTCTGCTCGGCTCCCAAGT 4767
Db 4139 GCCCAGGCTGTCTCAACCTCTGTGGCTCAAGCAATCTCTGCTCGGCTCCCAAGT 4198
QY 4768 GCTGGATTAAGATGTGAGCCACCGCATCCAGCCCAACCTCATTTATACCAATTAC 4827
Db 4199 GCTGGATTAAGATGTGAGCCACCGCATCCAGCCCAACCTCATTTATACCAATTAC 4258
QY 4828 CTGCCCCAGTACGTGGAGCTTTGCTTCACCCCGCTCTATCTGGAGAGAGAGGA 4887
Db 4259 CTGCCCCAGTACGTGGAGCTTTGCTTCACCCCGCTCTATCTGGAGAGAGAGGA 4318
QY 4888 TTATGTTATAGTCTGTACAGCAGTCCCAAGTTCAATTTTCTGCGCAAAAACTTCTT 4947
Db 4319 TTATGTTATAGTCTGTGTACAGCAGTCCCAAGTTCAATTTTCTGCGCAAAAACTTCTT 4378
QY 4948 CAAAAAATAATGTACTTCAATTTGATTAATGATTAATGATTAATGATTAATGATTAATG 5007
Db 4379 CAAAAAATAATGTACTTCAATTTGATTAATGATTAATGATTAATGATTAATGATTAATG 4438
QY 5008 CTGTTCACATGGCATTAAGAAAGAAATTTTATAGTCTCTAATAGGCTGTAGTCAAA 5067
Db 4439 CTGTTCACATGGCATTAAGAAAGAAATTTTATAGTCTCTAATAGGCTGTAGTCAAA 4498
QY 5068 GACCTCTTGAACCTTTCAGAGATAGATATTTAATGATCCCTTGGCGTTGCTAT 5127
Db 4499 GACCTCTTGAACCTTTCAGAGATAGATATTTAATGATCCCTTGGCGTTGCTAT 4558
QY 5128 GGCACCTTTCCTTCTGAAAAGTGTGTTCTGCCAGTACCTTGGCTTGTGAGCCGA 5187
Db 4559 GGCACCTTTCCTTCTGAAAAGTGTGTTCTGCCAGTACCTTGGCTTGTGAGCCGA 4618
QY 5188 GATGCTGACCTGATTAAGGGCCAAAGAGGGGTGGGTTCTTCCCTCAGTGAAGAG 5247
Db 4619 GATGCTGACCTGATTAAGGGCCAAAGAGGGGTGGGTTCTTCCCTCAGTGAAGAG 4678
QY 5248 CCTTATATTGAATTCATCTGTGTGAGGCTTGAAGCTTCTGATTTCTGCAATTTCCCAACTC 5307
Db 4679 CCTTATATTGAATTCATCTGTGTGAGGCTTGAAGCTTCTGATTTCTGCAATTTCCCAACTC 4738
QY 5308 CCAAGCCCTTTCAGAGAGGATAGGTGCTCTGATTCACACCAAGGTGGGATTTGGCCCTT 5367
Db 4739 CCAAGCCCTTTCAGAGAGGATAGGTGCTCTGATTCACACCAAGGTGGGATTTGGCCCTT- 4797
QY 5368 CTTAAGCTGGCTACTTCTTCAACATCACCGACATACCTGTGCTGCAAGACACACAGT 5427
Db 4798 CTTAAGCTGGCTACTTCTTCAACATCACCGACATACCTGTGCTGCAAGACACACAGT 4857
QY 5428 GCCATTTCTCTTAACCTGAGGGGCTCAAAACTCCGTGCAAGTGTGCGGCTCTGAGACCA 5487
Db 4858 GCCATTTCTCTTAACCTGAGGGGCTCAAAACTCCGTGCAAGTGTGCGGCTCTGAGACCA 4917
```


[illegible]

QY 2946 GGAATCTTCTTAAGGCGAAGACATTTGAAATGATGAATTCGAATCTGATACAGTCATGA 3005
|||||
Db 73621 GGAATCTTCTTAAGGCGAAGACATTTGAAATGATGAATTCGAATCTGATACAGTCATGA 73562
QY 3006 CAGCTCATGTGCTCTCTAACTTAGGCTGTGCGGTTAGCCAGCCCTGTATATAGAGAGAGA 3065
|||||
Db 73561 CAGCTCATGTGCTCTCTAACTTAGGCTGTGCGGTTAGCCAGCCCTGTATATAGAGAGAGA 73502
QY 3066 GGCCCTGAGTACCTAGCATAGGCTTTCAGCAAGCCCTGATTCAGAGTGTAAACAGAG 3125
|||||
Db 73501 GGCCCTGAGTACCTAGCATAGGCTTTCAGCAAGCCCTGATTCAGAGTGTAAACAGAG 73442
QY 3126 CTTCGCCCTCTTCAAGAGACAGAGTTCCATTCCAAAGAGCCCTACCTGAGGTCCTACTCTC 3185
|||||
Db 73441 CTTCGCCCTCTTCAAGAGACAGAGTTCCATTCCAAAGAGCCCTACCTGAGGTCCTACTCTC 73382
QY 3186 ACTGGGGTCCCCAGAGATGAAAGACAGCAATGTCCTTTTATATATATTTATTTGGTGTG 3245
|||||
Db 73381 ACTGGGGTCCCCAGAGATGAAAGACAGCAATGTCCTTTTATATATATTTATTTGGTGTG 73322
QY 3246 CTGTGTATTTAAGAGATCAAAATGTATACACCTAGCTCTTTTACCTGACTAGTAAT 3305
|||||
Db 73321 CTGTGTATTTAAGAGATCAAAATGTATACACCTAGCTCTTTTACCTGACTAGTAAT 73262
QY 3306 AACTCATACTACTGTGTTGGATGCTGGGTTGTGACTTCTACTGACCGCTAGATAACG 3365
|||||
Db 73361 AACTCATACTACTGTGTTGGATGCTGGGTTGTGACTTCTACTGACCGCTAGATAACG 73302
QY 3366 TGTGCGCTGCCCCAGGTGGGGAATATTTTCAATCTGCCCCAAGCAAGAAAGATGTG 3425
|||||
Db 73301 TGTGCGCTGCCCCAGGTGGGGAATATTTTCAATCTGCCCCAAGCAAGAAAGATGTG 73142
QY 3426 TGTGTTTGAAGAGATGACACATATCTGTTGATTAAGAGACTTCTGATTTCTAGT 3485
|||||
Db 73141 TGTGTTTGAAGAGATGACACATATCTGTTGATTAAGAGACTTCTGATTTCTAGT 73082
QY 3486 CGGTTCTGTGTTATCCCATTTGTGGAATTCATCTTGAATCCCATTTGCTATAGTCTAG 3545
|||||
Db 73081 CGGTTCTGTGTTATCCCATTTGTGGAATTCATCTTGAATCCCATTTGCTATAGTCTAG 73022
QY 3546 CAATTAAGAGAAATTTCTCAAGTTTCATGCGGTTCTCTTACCTGACAGCAATACCTTG 3605
|||||
Db 73021 CAATTAAGAGAAATTTCTCAAGTTTCATGCGGTTCTCTTACCTGACAGCAATACCTTG 72962
QY 3606 ACATTTAAAGAGAAATTTAGAGATATTCATCTCTAATAAATGTTTAAATATATACCA 3665
|||||
Db 72961 ACATTTAAAGAGAAATTTAGAGATATTCATCTCTAATAAATGTTTAAATATATACCA 72902
QY 3666 AACAGTGGCCCCCTGCAATTAGTTTCTGTGCGCACTGCAACCCATTTCTTGATGCTTAA 3725
|||||
Db 72901 AACAGTGGCCCCCTGCAATTAGTTTCTGTGCGCACTGCAACCCATTTCTTGATGCTTAA 72842
QY 3726 AAACAACACATTTAGCTATATGCTCTGGGATCAGAAATTCAGAAATGATGCTGATG 3785
|||||
Db 72841 AAACAACACATTTAGCTATATGCTCTGGGATCAGAAATTCAGAAATGATGCTGATG 72782
QY 3786 AAAATCAAGGTGTCAGAGAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGCTTCC 3845
|||||
Db 72781 AAAATCAAGGTGTCAGAGAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGCTTCC 72722
QY 3846 TTGCAATTTCAAGCTTTAGAGGCTGGCTGCAATTCAGAGGCTCAGTGGCTCAAGCT 3905
|||||
Db 72721 TTGCAATTTCAAGCTTTAGAGGCTGGCTGCAATTCAGAGGCTCAGTGGCTCAAGCT 72662
QY 3906 TTTTCTCATGGCATCTGACAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGCTTCC 3965
|||||
Db 72661 TTTTCTCATGGCATCTGACAGCTGTGCTCTTCTGAAGGCTTAGGAGAGAGCCGCTTCC 72602
QY 3966 CCACACAGAGATCCAGAGATATCTCTCAATCTAAAGATCCCTTATCATCTGGAAGAGC 4025
|||||
Db 72601 CCACACAGAGATCCAGAGATATCTCTCAATCTAAAGATCCCTTATCATCTGGAAGAGC 72542

QY 4026 CTTTGGCATGCAAGACACATATGCCACAGGTGGGGATTAGGACAGAGACATCTTTGGG 4085
|||||
Db 72541 CTTTGGCATGCAAGACACATATGCCACAGGTGGGGATTAGGACAGAGACATCTTTGGG 72482
QY 4086 TGCCTTATTTCTGCTTCCACACCTTCTGCACTGTGCTCCACAGAGAGGCTACAAA 4145
|||||
Db 72481 TGCCTTATTTCTGCTTCCACACCTTCTGCACTGTGCTCCACAGAGAGGCTACAAA 72422
QY 4146 TGATCTGGCCACAGAGGATTTTGTATGCTTGGGACTCTAACCTT-AAAAAACC 4204
|||||
Db 72421 TGATCTGGCCACAGAGGATTTTGTATGCTTGGGACTCTAACCTT-AAAAAACC 72362
QY 4205 CAGATCAGAAATCTGGCCATGCTGGGCTCACATTTCTACCTAGCAACAATGCTGGA 4264
|||||
Db 72361 CAGATCAGAAATCTGGCCATGCTGGGCTCACATTTCTACCTAGCAACAATGCTGGA 72302
QY 4265 GCTGGGACACAGCTCTGCTTTAGAAAGGGGTGCATTCACAGGCTCACACAGCCAC 4324
|||||
Db 72301 GCTGGGACACAGCTCTGCTTTAGAAAGGGGTGCATTCACAGGCTCACACAGCCAC 72242
QY 4325 ACTAGCCCTATCAGCTTCCACAAATGAGGCTAAGTGTGTTCTACTGATCAATGCCCC 4384
|||||
Db 72241 ACTAGCCCTATCAGCTTCCACAAATGAGGCTAAGTGTGTTCTACTGATCAATGCCCC 72182
QY 4385 TGCAGTTGCATTTATTTGTAATGAAAAAGAAAGACTGGATTAATCTTAATCAGTGA 4444
|||||
Db 72181 TGCAGTTGCATTTATTTGTAATGAAAAAGAAAGACTGGATTAATCTTAATCAGTGA 72122
QY 4445 TAGACATGAGACCAATGTGCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT 4504
|||||
Db 72121 TAGACATGAGACCAATGTGCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT 72062
QY 4505 CTTTTTTTTTAAATGAGACAGATCTATCTGTTGCTAGGCTGAGGTGAGTGGC 4564
|||||
Db 72061 CTTTTTTTTTAAATGAGACAGATCTATCTGTTGCTAGGCTGAGGTGAGTGGC 72002
QY 4565 GCATCTGGGCTCACTGCAACCTCTGCTCTGGGCTCAAGCAATTTCTCCACTCAGCC 4624
|||||
Db 72001 GCATCTGGGCTCACTGCAACCTCTGCTCTGGGCTCAAGCAATTTCTCCACTCAGCC 71942
QY 4625 TCCCAATAGCTGGGATCAGTGGCAACCAACCATGCCAGCTAAATTTTGTATTTTGG 4684
|||||
Db 71941 TCCCAATAGCTGGGATCAGTGGCAACCAACCATGCCAGCTAAATTTTGTATTTTGG 71882
QY 4685 TAGAGACAGGGTTTACCATGTTGCCCCAGGCTGTCAACCTCTGSGCTCAAGCAATC 4744
|||||
Db 71881 TAGAGACAGGGTTTACCATGTTGCCCCAGGCTGTCAACCTCTGSGCTCAAGCAATC 71822
QY 4745 CTCTGGCTGGGCTCCCAAAAGTGTGGGANTTACAGATGTGAGCCACCGCATCCAGCCC 4804
|||||
Db 71821 CTCTGGCTGGGCTCCCAAAAGTGTGGGANTTACAGATGTGAGCCACCGCATCCAGCCC 71762
QY 4805 ACACCCCTATTTATACCAATTTACCTGCGCAGTAAGTGTGCACTTTTCTTCTCAACCT 4864
|||||
Db 71761 ACACCCCTATTTATACCAATTTACCTGCGCAGTAAGTGTGCACTTTTCTTCTCAACCT 71702
QY 4865 GCTCTGATCTGGAAGAGAGGATTAATATAGCTTTGTCAGACAGCTCCCAAGTCAAT 4924
|||||
Db 71701 GCTCTGATCTGGAAGAGAGGATTAATATAGCTTTGTCAGACAGCTCCCAAGTCAAT 71642
QY 4925 ATTTCTGGGCAAAAATCTTCTTCAAAAATTAATATGACTTCAATTTGATTAATGATTC 4984
|||||
Db 71641 ATTTCTGGGCAAAAATCTTCTTCAAAAATTAATATGACTTCAATTTGATTAATGATTC 71582
QY 4985 ACCCTGGAATGACCGCCCTCAACTTTTTCACATGAGATTAATGAAGAAATTTATAGT 5044
|||||
Db 71581 ACCCTGGAATGACCGCCCTCAACTTTTTCACATGAGATTAATGAAGAAATTTATAGT 71522
QY 5045 CTCTTAATGGGCTGTACTGCAAGACCTCTTGAACATTTTCCAGAGATAGGATATTTAA 5104
|||||
Db 71521 CTCTTAATGGGCTGTACTGCAAGACCTCTTGAACATTTTCCAGAGATAGGATATTTAA 71462
QY 5105 GTCATGGCTTGGGCTTGGCTATGGACCTTTCCCTTCTGAAGTGTGTTCTGCCCCAG 5164

```

|||||
Db 71461 GTCATGCCCTTGGCGTTCCTATGGCACTTCCCTTCAAGGTGCTGTTCCGTGCCAG 71402
5165 TGACCCCTTGGCTTGTGAGCCGAGATGCTGACCTGATTAAGGCCAAAGAGGGCTGC 5224
71401 TGACCCCTTGGCTTGTGAGCCGAGATGCTGACCTGATTAAGGCCAAAGAGGGCTGC 71342
5225 GGCTTCCTTCCCTCAGTGAAGAGCCCTTATTTGAATTCAGTGTGAGGCCCTAGGCCCTC 5284
71341 GGCTTCCTTCCCTCAGTGAAGAGCCCTTATTTGAATTCAGTGTGAGGCCCTAGGCCCTC 71282
5285 CATTTCGACATTTCCCAACCTCCCAAGCCCTTCAAGAGAGAGTGGTCCCTGCAATTC 5344
71281 CATTTCGACATTTCCCAACCTCCCAAGCCCTTCAAGAGAGAGTGGTCCCTGCAATTC 71222
5345 CACCAAGGTGGAGATTGGCTTCCCTTAAGCTGCTACTGTCAACCATCACGACATCACT 5404
71221 CACCAAGGTGGAGATTGGCTTCCCTTAAGCTGCTACTGTCAACCATCACGACATCACT 71162
5405 GTTGGCTGCAAGGACACACAGTGGCCATTTCTTCAACAGAGGGCTCAAAACTCCTGGA 5464
71161 GTTGGCTGCAAGGACACACAGTGGCCATTTCTTCAACAGAGGGCTCAAAACTCCTGGA 71102
5465 CAAATTGCTGGCTCTGAGACAGATATTTCTGGAGCTGTGCTCAAGTGAAGGGGCCAG 5524
71101 CAAATTGCTGGCTCTGAGACAGATATTTCTGGAGCTGTGCTCAAGTGAAGGGGCCAG 71042
5525 CCTGAGAACCCCTGGCTCTTTTCTTAAAGCCAGGCCCACTTACATTAACATTTGAG 5584
71041 CCTGAGAACCCCTGGCTCTTTTCTTAAAGCCAGGCCCACTTACATTAACATTTGAG 70982
5585 GGTCACGTGGAACAGTGAAGTGGCATTTGTTGAAGCTTACTGCATGTCACACCCACTGCTC 5644
70981 GGTCACGTGGAACAGTGAAGTGGCATTTGTTGAAGCTTACTGCATGTCACACCCACTGCTC 70922
5645 ATCCAGCTGGTCTGCAATGCTGACGAGAGGACGACGATGAGAGCTGCTCTAATG 5704
70921 ATCCAGCTGGTCTGCAATGCTGACGAGAGGACGACGATGAGAGCTGCTCTAATG 70862
5705 CTGTGCTATTCACAGAGAGGAAAGGCTCAAGAGAGTCAACTGGAACAAGCACAG 5764
70861 CTGTGCTATTCACAGAGAGGAAAGGCTCAAGAGAGTCAACTGGAACAAGCACAG 70802
5765 CCCACGGACATGGCCCTTGTAAAGTTAGACACAGCTGTGTGATCTGCAGTCTT 5824
70801 CCCACGGACATGGCCCTTGTAAAGTTAGACACAGCTGTGTGATCTGCAGTCTT 70742
5825 CACTGGAATTAATTTATTCATTCAGATCACTTTTAAAGTGGCATTTTATTCATTTCTGT 5884
70741 CACTGGAATTAATTTATTCATTCAGATCACTTTTAAAGTGGCATTTTATTCATTTCTGT 70682
5885 GCTTTAATTAACAATGTAACAAAAACAAGATCAAGCTGTTTAAAGTCTTGGCTTAC 5944
70681 GCTTTAATTAACAATGTAACAAAAACAAGATCAAGCTGTTTAAAGTCTTGGCTTAC 70622
5945 TTGTCCCTGGTTCAGTGAAGGCCCGGTTTCCAGTGTGACTGTGATGAGAGCTCAGCA 6004
70621 TTGTCCCTGGTTCAGTGAAGGCCCGGTTTCCAGTGTGACTGTGATGAGAGCTCAGCA 70562
6005 TGGGCTCAGAGATGCTGCTTAAATTTGTGATGATACAGAAAGCCAGCTTTGGGATAC 6064
70561 TGGGCTCAGAGATGCTGCTTAAATTTGTGATGATACAGAAAGCCAGCTTTGGGATAC 70502
6065 AAGTCTTTCCTCTCATTTGATGATGCGTGCATGTGTGAAGCAGATGTTTTCGCGAA 6124
70501 AAGTCTTTCCTCTCATTTGATGATGCGTGCATGTGTGAAGCAGATGTTTTCGCGAA 70442
6125 ATTAATAATAATAGTCTTGAGTCTCGCA 6153
70441 ATTAATAATAATAGTCTTGAGTCTCGCA 70413

```

RESULT 6

```

AC104165/
LOCUS      AC104165      200956 bp      DNA      linear      PRI 27-JUN-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-68104, complete sequence.
ACCESSION AC104165 AC068625
VERSION    AC104165.2 GI:121617715
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 200956)
            Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
            Direct Submission
            Unpublished
            2 (bases 1 to 200956)
            Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
            Direct Submission
            Submitted (05-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
            3 (bases 1 to 200956)
            Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
            Direct Submission
            Submitted (27-JUN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
            On Jun 27, 2002 this sequence version replaced gi:17352429.
            ----- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: WUGSC
            ----- Project Information
            Center project name: chr-3
            Center clone name: RP11-68104 (bc0528)
            ----- Summary Statistics
            Sequencing vector: unknown; 42% of reads
            Sequencing vector: plasmid; 58% of reads
            Chemistry: Dye-terminator ET; 80% of reads
            Chemistry: Dye-terminator Big Dye; 20% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 200732 bases at least Q40
            Consensus quality: 200909 bases at least Q30
            Consensus quality: 200947 bases at least Q20
            Insert size: 200956; sum-of-ctrls
            Quality coverage: 10.4x in Q20 bases; sum-of-ctrls
            -----
            Overlapping Sequences:
            5': RP11-348P10 (UWGC:bc0377) AC124045
            3': RP11-578F5 (UWGC:bc0477) AC105902, 95352-bp overlap
            -----
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            -----
            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            -----
            Sequence Validation:

```

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII		EcoRI		HindIII	
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
5328	5288	8696	8571	3534	3560
2067	2057	6	<800	6382	6318
6846	6934	6068	6154	512	<800
144	<800	11400	11207	449	<800
1459	1439	96	<800	5736	5646
70	<800	4100	4106	2978	3031
3690	3725	779	760	8146	8133
102	<800	1221	1170	4465	4477
511	<800	278	<800	2239	2257
1903	1919	8458	8571	45	<800
5124	5028	12014	11879	4368	4477
367	<800	645	<800	286	<800
7100	6934	1177	1170	4633	4784
5890	5889	153	<800	1462	1405
1814	1816	1993	1932	2058	2110
125	<800	3153	3186	745	<800
43	<800	132	<800	1321	1243
117	<800	4939	4965	3528	3560
1955	1919	466	<800	1640	1662
3477	3529	1311	1279	8779	8765
4899	5028	3183	3186	5089	4991
774	799	16952	16678	2571	2631
10749	10616	1629	1629	2440	2446
1620	1602	7261	7427	1682	1662
8868	8810	7837	7939	1660	1662
5884	5889	3940	4106	4912	4991
1521	1529	756	760	2470	2446
771	799	2395	2496	8892	8765
11706	11675	5310	5399	1257	1243

369	<800	21139	21498	10114	9869
5177	5288	4082	4106	14019	14339
2916	2950	7866	7939	3580	3560
137	<800	7411	7427	3022	3031
2406	2347	3378	3366	8296	8133
3857	3725	7272	7427	3319	3311
5249	5288	3059	3186	1694	1662
5343	5288	8214	8571	1747	1662
6819	6934	5836	5849	2370	2446
2335	2347	5394	5399	4817	4784
569	<800	419	<800	1139	1073
135	<800	3260	3186	3946	3909
3825	3725	2082	2097	12931	12902
6741	6934	8568	8571	2109	2110
1472	1439	1681	1629	364	<800
3707	3725	741	760	78	<800
703	<800	2905	3005	254	<800
207	<800			6129	6318
5047	5028			1674	1662
1134	1133			1250	1243
1070	1069			4768	4784
4665	4675			1083	1073
1121	1133			1023	1073
6920	6934			2602	2631
3317	3264			2449	2446
1683	1697			1249	1243
3525	3529			7228	7182

Query Match 61.2%; Score 3769.8; DB 9; Length 200956;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 3791; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 2346 TTGCTGTGTGAAAAAGAGAAAAAGAACACAAAGGCCCCCTGTGGTATCTACAA 2405
 Db 179926 TTCCATTTTTCCTCAATATGAAAAAGAACAAAGGCCCCCTGTGGTATCTACAA 179867
 QY 2406 TTGGCAACATCAATCTAGATGCGGAGGCAAGCCAAAAAGTTTCAGAAAGGCGCAAGGA 2465
 Db 179866 TGACAAACATCAATCTAGATGCGGAGGCAAGCCAAAAAGTTTCAGAAAGGCGCAAGGA 179807
 QY 2466 CAATGACTCCCATGTGTATGCAAGTCAATGAGAGACACCATGGTATATGGCATCTGTACA 2525
 Db 179806 CAATGACTCCCATGTGTATGCAAGTCAATGAGAGACACCATGGTATATGGCATCTGTACA 179747
 QY 2526 GGATTCCAGGGGCTCTCTCTGACAGCCAGAGTGGACACCTACCGGCGGTTCCAGGGGCAC 2585

D	b	179746	GGATTCCAGCGGCTCTCTTCCGTGCAAGCAGAGGTGGACACTTACCGGCGCTTCCAGGGCAC	179687
O	y	2586	CATGGGGCTCTGTCTCTCCCTCCCAACCACACATATGCTCCAGGGGCCCACTGCAAGTT	2645
D	b	179686	CATGGGGTCTGTCTCTCTCCCTCCCAACCACATATGCTCCAGGGGCCCACTGCAAGTT	179622
O	y	2646	GGCCACTGAGGAGCCACTCTCGCTCCCTCTGAGTCTGAGAGTGAACCGTACACTT	2705
D	b	179626	GGCCACTGAGGAGCCACTCTCGCTCCCTCTGAGTCTGAGAGTGAACCGTACACTT	179567
O	y	2706	CTCCCATCCCAACATGGGGATGTATACGAGAAAGACACAGACTTTCCTTACTGAACAC	2765
D	b	179566	CTCCCATCCCAACATGGGGATGTATACGAGAAAGACACAGACTTTCCTTACTGAACAC	179507
O	y	2766	TCAGAGAGCCCATGTGAGCCAGAGATAATCTGATTCATTCAGACGGCTTGTGAGATTTC	2825
D	b	179506	TCAGAGAGCCCATGTGAGAGCAGAGATAATCTGATTCATTCAGACGGCTTGTGAGATTTC	179447
O	y	2826	ATAAAGCAGGGCACTGAGACACCCGTCGTTCTTAACAGAAATCCTAAAGAAAGAGA	2885
D	b	179446	ATAAAGCAGGGCACTGAGACACCCGTCGTTCTTAACAGAAATCCTAAAGAAAGAGA	179387
O	y	2886	ATTATACGAAGAAACAGACAGAGAGTTTCTGGAACACCGCACTTACATTGCTAGT	2945
D	b	179386	ATTATACGAAGAAACAGACAGAGAGTTTCTGGAACACCGCACTTACATTGCTAGT	179327
O	y	2946	GGACTCATTTCTTAAGGGCAGACATTTGAAATGATGAATTCGAACTGATACAGTCTGA	3005
D	b	179326	GGACTCATTTCTTAAGGGCAGACATTTGAAATGATGAATTCGAACTGATACAGTCTGA	179267
O	y	3006	CAGCTCATGTCTCTCTCAACTTAGGCTGTGCGTTAGCCACCCGTATAGAGAGAGAGA	3065
D	b	179266	CAGCTCATGTCTCTCTCAACTTAGGCTGTGCGTTAGCCACCCGTATAGAGAGAGAGA	179207
O	y	3066	GGCCTGATTCACCTTAGATTAGGGTTGCAAGCAAGCCCTGATTCAGAGTGTTAAACAGAG	3125
D	b	179206	GGCCTGATTCACCTTAGATTAGGGTTGCAAGCAAGCCCTGATTCAGAGTGTTAAACAGAG	179147
O	y	3126	CTTGCCCTCTTCAGAGACAAAGTTCCAAATTCAGAGAGCCCTGAGAGGCCCTACTC	3185
D	b	179146	CTTGCCCTCTTCAGAGACAAAGTTCCAAATTCAGAGAGCCCTGAGAGGCCCTACTC	179087
O	y	3186	ACTGGGGTCCCAAGAGTAAAGACAAATGTGCCCTTTATATATATATATTTGGTGGTC	3245
D	b	179086	ACTGGGGTCCCAAGAGTAAAGACAAATGTGCCCTTTATATATATATATTTGGTGGTC	179027
O	y	3246	CTGTGTATTTTAAAGAGATCAAAATGTATTAACCACTAGCTCTTTACCTGACTTAGTAAT	3305
D	b	179026	CTGTGTATTTTAAAGAGATCAAAATGTATTAACCACTAGCTCTTTACCTGACTTAGTAAT	178967
O	y	3306	AACTCATACTTAAGTGGTTGGATGCGTGGGGTTGTGACTCTACAGACGGGTAGTAAAG	3365
D	b	178966	AACTCATACTTAAGTGGTTGGATGCGTGGGGTTGTGACTCTACAGACGGGTAGTAAAG	178907
O	y	3366	TGTGCTGTGCCCAAGGTGTGGGAATTAATTACAAATCTGTCCAACAGAAAGAAATGTG	3425
D	b	178906	TGTGCTGTGCCCAAGGTGTGGGAATTAATTACAAATCTGTCCAACAGAAAGAAATGTG	178847
O	y	3426	TGTGTTTAGACAGCATTGACACATATCTGCTTTATTAAGAGACTTCTGATTCTTAGGT	3485
D	b	178846	TGTGTTTAGACAGCATTGACACATATCTGCTTTATTAAGAGACTTCTGATTCTTAGGT	178787
O	y	3486	CGGTTCCGGGTTATCCCATCTTGGAAATTCATCTTGAATCCCATGTGCTATAGTCTTAG	3545
D	b	178786	CGGTTCCGGGTTATCCCATCTTGGAAATTCATCTTGAATCCCATGTGCTATAGTCTTAG	178727
O	y	3546	CAATTAAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGCTCCAGCAATACTTTG	3605
D	b	178726	CAATTAAGAAATTTCTCAAGTTTCCATGTGCGGTTCTCTAGCTCCAGCAATACTTTG	178667
O	y	3606	ACATTTAAAGAAATTTAGAGATATTTCTCATCTCTTAAAAATGTTTAAATATATACCA	3665

D	b	178666	ACATTAAAGAGAAATTAGAGAAATATTCTCATCTCTAAAAATGTTTAATATATACCA	178607
Q	y	3666	AACACTGGCCCCCTGCATATTGTTTCTGTGTCACATGCACACCACTTACTTGTTAGCTTAA	3725
D	b	178606	AACACTGGCCCCCTGCATATTGTTTCTGTGTCACATGCACACCTTACTTGTTAGCTTAA	178547
Q	y	3726	AAACACACATTAGCTATTAGTCTGTGGGATCAGAAATTCAAAATGATGTCCCTGAATG	3785
D	b	178546	AAACACACATTAGCTATTAGTCTGTGGGATCAGAAATTCAAAATGATGTCCCTGAATG	178487
Q	y	3786	AAAATCAAGGTGTCAGCAGACTGTGGCTCTCTGAAAGGCTCTGGGAGAAAGCCGGTTCC	3845
D	b	178486	AAAATCAAGGTGTCAGCAGACTGTGGCTCTCTGAAAGGCTCTGGGAGAAAGCCGGTTCC	178427
Q	y	3846	TTGCATTTCAAAGCTCTTAGAGAGCTGCTGCATTTCCAGAGCTCCAGTGGCTCAAGCT	3905
D	b	178426	TTGCATTTCAAAGCTCTTAGAGAGCTGCTGCATTTCCAGAGCTCCAGTGGCTCAAGCT	178367
Q	y	3906	TTTTCTACATGGCATCTACTGTGACACTGTGGCCCTCCCACTTCCCTTTGACTTACAAGC	3965
D	b	178366	TTTTCTACATGGCATCTACTGTGACACTGTGGCCCTCCCACTTCCCTTTGACTTACAAGC	178307
Q	y	3966	CCACAGAGAAAGTCAGGATTAATCTCTCCATCTTAAAGTCCCTTATCATCTCTGGAAGAGC	4025
D	b	178306	CCACAGAGAAAGTCAGGATTAATCTCTCCATCTTAAAGTCCCTTATCATCTCTGGAAGAGC	178247
Q	y	4026	CTTTTGCCATGCAAGACAACTATAGCCACAGGTGGGAGATTAGACACAGACATCTTTGGG	4085
D	b	178246	CTTTTGCCATGCAAGACAACTATAGCCACAGGTGGGAGATTAGACACAGACATCTTTGGG	178187
Q	y	4086	TGCTGTATTCTGCTCCACACACTTCTCTGCACTGACTCCACAGAGAGGCTACAAA	4145
D	b	178186	TGCTGTATTCTGCTCCACACACTTCTCTGCACTGACTCCACAGAGAGGCTACAAA	178127
Q	y	4146	TGATCTGGCCGACAGAGGATGTTTTGTTTAGTCTGCGAGCTCTAACACTT-AAAAAAACC	4204
D	b	178126	TGATCTGGCCGACAGAGGATGTTTTGTTTAGTCTGCGAGCTCTAACACTT-AAAAAAACC	178067
Q	y	4206	CAGATCAGAAGATCTGGGCATGTGGGGCTCACATTTCACACTTACGAACAATGGCTGGA	4264
D	b	178066	CAGATCAGAAGATCTGGGCATGTGGGGCTCACATTTCACACTTACGAACAATGGCTGGA	178007
Q	y	4265	GCTGGGCAACCACTCTGCTTTTGAAGAGGGTGTCCACTTCACAGTCCACACAGCCAC	4324
D	b	178006	GCTGGGCAACCACTCTGCTTTTGAAGAGGGTGTCCACTTCACAGTCCACACAGCCAC	1779447
Q	y	4325	ACTAGGCCCTATCACTTCTCCCAACAATGAGGCTTAAGTGTGTTCTACTGATCATGTCCCC	4384
D	b	177946	ACTAGGCCCTATCACTTCTCCCAACAATGAGGCTTAAGTGTGTTCTACTGATCATGTCCCC	177887
Q	y	4385	TGCAGGTTCGATTATTGTTAATGAAAAAGAAAGATGGGATTAATCTCTAATCAGGTGAG	4444
D	b	177886	TGCAGGTTCGATTATTGTTAATGAAAAAGAAAGATGGGATTAATCTCTAATCAGGTGAG	177827
Q	y	4445	TAGACCATGAGACCAATGTGTCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT	4504
D	b	177826	TAGACCATGAGACCAATGTGTCTCACATTAACCTTTTCTTTTCTTTTCTTTTCTTTT	177767
Q	y	4505	CTTTTCTTTTCTTTTAAATGTAGACAGAGATCTCAATCTGTGCTCAGCTGAGAGTGCAGTGGC	4564
D	b	177766	CTTTTCTTTTCTTTTAAATGTAGACAGAGATCTCAATCTGTGCTCAGCTGAGAGTGCAGTGGC	177707
Q	y	4565	GCAATCTGGGCTCACTGTGAAACCTGCTCCTGGGGCTCAAGCAATTTCTCCACCTCAGCC	4624
D	b	177706	GCAATCTGGGCTCACTGTGAAACCTGCTCCTGGGGCTCAAGCAATTTCTCCACCTCAGCC	177647
Q	y	4625	TCCCAATAGCTGGGATCACTGGCAAAACCAACCATGCCCCAGCTAATTTGTATTTTTT	4684
D	b	177646	TCCCAATAGCTGGGATCACTGGCAAAACCAACCATGCCCCAGCTAATTTGTATTTTTT	177587
Q	y	4685	TAGAGACAGGGTTTACCATGTGTCAGAGGCTGTCTCAACTCTGTGGCTCAAGCAATC	4744
D	b	177586	TAGAGACAGGGTTTACCATGTGTCAGAGGCTGTCTCAACTCTGTGGCTCAAGCAATC	177527


```
/translation="MAGLNCGVSTALLGLVLLGAARLRPGAEFETALPRESNTVLI
KLPPEVLAKPCYIYISKRHTIMLSIKSGERIVTFSSQSPENHNHVIETOKRIDMSG
PCPPEVUOLPSTSLPLTLPNTFRFMDVNAKHSIGLELPSLRLQIGGESCPGVT
HSISGIDATVVRIGTFCSNGTISRKIQKEGVKMLHPWEPRAVSGSLANRSIK
RLCIIIEVFEBSGATLMSANPEGPPEDELTMTQFVVPAPHRAVSELTMSNCER
KEEREYIIPGSTNPFEVKLEDKOPGMAGNENSLSGCDDADOSPGLRLFOFVU
OHPOENSKYIVDLSENRAVSLTIEPRVQSRKRFVGPFCLESPRCSLSLTJSG
SKHISPLCDLTLRUMNVEKTIESTDRKORXSYSLQVPSDLHLHVEHDEFSMKL
LVPKRLSLVLPVPAKLDQHTHEKPCNTSFSLVASLPSDLIYGSFCPGSIOQ
VKONISVLRFPSEFQESRQSLJVFIPYKEEGVTFIPDTSKVYLRTPMMDR
GLPSTISVSNISVPRDQVACLTFEKRSGVVCQGRAMIIQOERBAEELFSD
VLKPSFHHHSKMNINSKSPSGKOLDFEVLTLTPRVDLVLIAVGGVLLS
ALGLICCVKKKKKKNKPAVGIYNGINMTNPPROPKFOKGRDNDNDAVIEDT
MYGHLDDSSSFLQPEVDITRRPFGMGVCPPEPPTICSAAPLAKLATEEPPRSP
PESBEPTFFSHPNMGDVSKDTPILPLSTDEMEPAE"
BASE COUNT      789 a. 922 c 810 g 703 t
ORIGIN
Query Match      51.9%: Score 3201.6; DB 9; Length 3224;
Beet Local Similarity 99.9%: Pred. No. 0;
Matches 3204; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 173 GCGCGAGGTGAGTAGAGCCAGGCGGAGCGAGCTGCGCGGGCTTGGCGCGCTGGGCGC 232
    |||||
Db 1 GCGCGAGGTGAGTAGAGCCAGGCGGAGCGAGCTGCGCGGGCTTGGCGCGCTGGGCGC 60
QY 233 GCGCGTCCCGACCGTGTGTTTCCCGACCGAGCGCGTCCCGGAGTCATGGCGCGC 292
    |||||
Db 61 GCGCGTCCCGACCGTGTGTTTCCCGACCGAGCGCGTCCCGGAGTCATGGCGCGC 120
QY 293 TGAAGTGGGGGTCTGTATCGACATGTAAGGGGTGTGCTGCTGGGTGGGGGCGGCTGC 352
    |||||
Db 121 TGAAGTGGGGGTCTGTATCGACATGTAAGGGGTGTGCTGCTGGGTGGGGGCGGCTGC 180
QY 353 GCGCGGCGGAGAGCTTTTGAGATTGCTCTGCCAGAGAAAGCAATTACAGTTCTCA 412
    |||||
Db 181 GCGCGGCGGAGAGCTTTTGAGATTGCTCTGCCAGAGAAAGCAATTACAGTTCTCA 240
QY 413 TAAAGTGGGGAGCCCGACTGTGCTGCGACAGAAAGCAATTACAGTTCTCA 472
    |||||
Db 241 TAAAGTGGGGAGCCCGACTGTGCTGCGACAAACCCGTGTACATCGTCATTTCTAAAGAGC 300
QY 473 ATATAACATGTTGTTCATCAAGTGTGAGAAAGAAATAGTCTTACCTTACCTGCGAGA 532
    |||||
Db 301 ATATAACATGTTGTTCATCAAGTGTGAGAAAGAAATAGTCTTACCTTACCTGCGAGA 360
QY 533 GTCTGAGAGATCACTTGTCTATAGAGATCAGAAAAATATGACTATGTACAGGCCAT 592
    |||||
Db 361 GTCTGAGAGATCACTTGTCTATAGAGATCAGAAAAATATGACTATGTACAGGCCAT 420
QY 593 GTCTTGTGGGAGGTTCAAGCTTCAAGCCTCGACATGTTGTTGCTTACCTTACAGAA 652
    |||||
Db 421 GTCTTGTGGGAGGTTCAAGCTTCAAGCCTCGACATGTTGTTGCTTACCTTACAGAA 480
QY 653 GTTTCATGCGGATGTCAAGGTCATTAAGAGATGCGTTTAAAGTCGAGTTTTCATCC 712
    |||||
Db 481 GTTTCATGCGGATGTCAAGGTCATTAAGAGATGCGTTTAAAGTCGAGTTTTCATCC 540
QY 713 CTGCGCTGAGGAGATCGGTCCGGGTGAGAGCTGCCAGAGGAGTCACTCATCTCATCA 772
    |||||
Db 541 CTGCGCTGAGGAGATCGGTCCGGGTGAGAGCTGCCAGAGGAGTCACTCATCTCATCA 600
QY 773 GCGGCGGAATCGATGCGCACCGTGTGAGATCGGAACCTTGTGACAGCAATGCGACTGTG 832
    |||||
Db 601 GCGGCGGAATCGATGCGCACCGTGTGAGATCGGAACCTTGTGACAGCAATGCGACTGTG 660
QY 833 CCGGAGTCAAGATGCAAGAGAGTGAAGATGCGTTACACGCCCATGCTTCCAGCCCA 892
    |||||
Db 661 CCGGAGTCAAGATGCAAGAGAGTGAAGATGCGTTACACGCCCATGCTTCCAGCCCA 720
QY 893 GAAATGTCTCGGCTTCAGCATTTGCAAAACCGCTCATCTATAAAACCTGTGTGATCATG 952
    |||||
Db 721 GAAATGTCTCGGCTTCAGCATTTGCAAAACCGCTCATCTATAAAACCTGTGTGATCATG 780
```


OY	2033	AGGAGGCGCTGCTCACTTTCTTTAAAGAGGGAGCGCGGTGCTGCGCAGACAGGGCGC	20922
Dd	1861	AGGTGGCGCTGCTCACTTTCTTTAAAGAGCGAGCGCGGTGCTGCGCAGACAGGGCGC	19200
OY	2093	CATTTCATGATCATCCAGAGACAGCGGAGCCGGGCTGAGAGATCTTTCAGCCGTGAGAGG	21522
Dd	1921	CATTTCATGATCATCCAGAGACAGCGGAGCCGGGCTGAGAGATCTTTCAGCCGTGAGAGG	19800
OY	2153	ATTGTCGTCCCAAGCCAAAGTTCACACATTCACAGCTTTCGGGTCAACATCTCTAACTGCA	22122
Dd	1981	ATTGTCGTCCCAAGCCAAAGTTCACACATTCACAGCTTTCGGGTCAACATCTCTAACTGCA	20400
OY	2213	GCCCCAGAGGGGGAAACACACTGAGCTGTCTTCTCGGTGACACTTACCTTACCCAGAGAGTG	22722
Dd	2041	GCCCCAGAGGGGGAAACACACTGAGCTGTCTTCTCGGTGACACTTACCTTACCCAGAGAGTG	21000
OY	2273	TGGACTTGACTGTCATCTCATTCAGCAGCGGTGGAGAGTGGAGTCTTACTGCTGTGCGCC	23322
Dd	2101	TGGACTTGACTGTCATCTCATTCAGCAGCGGTGGAGAGTGGAGTCTTACTGCTGTGCGCC	21600
OY	2333	TGCGGCTCATCATTTGCTGTGTAAAAAGAAAAAGAAAGCAACAAAGGGCCCCGCTG	23922
Dd	2161	TGCGGCTCATCATTTGCTGTGTAAAAAGAAAAAGAAAGCAACAAAGGGCCCCGCTG	22200
OY	2393	TGGGATCTCAAAAGGCAATCAATACATGATGTCGGCAGGAGCAAAAAAGTTTCAGA	24522
Dd	2221	TGGGATCTCAAAAGGCAATCAATACATGATGTCGGCAGGAGCAAAAAAGTTTCAGA	22800
OY	2453	AAGGGCGAAAGGACAATGACTCCCATGTGATTCATCGAGAGACACATGATATG	25122
Dd	2281	AAGGGCGAAAGGACAATGACTCCCATGTGATTCATCGAGAGACACATGATATG	23400
OY	2513	GGCATCTGCTACAGAGATTCCAGCGGCTCCTTCTGCAAGCCAGAGGTGGACACTACCGGC	25722
Dd	2341	GGCATCTGCTACAGAGATTCCAGCGGCTCCTTCTGCAAGCCAGAGGTGGACACTACCGGC	24000
OY	2573	GGTTCACAGGGACCATGSGGGGTCTGTCTCTCCGCCACCAACATATGCTCCAGGGGCC	26322
Dd	2401	GGTTCACAGGGACCATGSGGGGTCTGTCTCTCCGCCACCAACATATGCTCCAGGGGCC	24600
OY	2633	CAACTGCAAAAGTTGGCCACTGAGAGAGCCACCTTCGCTCCCTCCTGAGTGTGAGATG	26922
Dd	2461	CAACTGCAAAAGTTGGCCACTGAGAGAGCCACCTTCGCTCCCTCCTGAGTGTGAGATG	25200
OY	2693	AACCGTACACCTTCTCCATCCCAACATGGGGATGTAAAGCAGCAAGACACACATTC	27522
Dd	2521	AACCGTACACCTTCTCCATCCCAACATGGGGATGTAAAGCAGCAAGACACACATTC	25800
OY	2753	CGTTACTGACACACACAGAGCCCATGAGCGACACAAATATCTGATCATCTCCAGAGCG	28122
Dd	2581	CGTTACTGACACACACAGAGCCCATGAGCGACACAAATATCTGATCATCTCCAGAGCG	26400
OY	2813	TTTCTCTAGATTTCATTAAGCAGGGGCACTGAGACCCGCTCGGTTCCTAACACAGAATTC	28722
Dd	2641	TTTCTCTAGATTTCATTAAGCAGGGGCACTGAGACCCGCTCGGTTCCTAACACAGAATTC	27000
OY	2873	CTAAAGAAGAGGAATTTCACAGAGGAACGACGAGGTTTTCTGAGACACCGCAACTT	29322
Dd	2701	CTAAAGAAGAGGAATTTCACAGAGGAACGACGAGGTTTTCTGAGACACCGCAACTT	27600
OY	2933	CACATTCGCTCAGTGGACTCATTTAAAGGCAAGACATTTAAAAATGATGAATTCATATCG	29922
Dd	2761	CACATTCGCTCAGTGGACTCATTTAAAGGCAAGACATTTAAAAATGATGAATTCATATCG	28200
OY	2993	GATACAGTCATGACAGTCTCATGTGCTCCTCAACTTATGGGTGAGCGGTTTACGACGCTGA	30522
Dd	2821	GATACAGTCATGACAGTCTCATGTGCTCCTCAACTTATGGGTGAGCGGTTTACGACGCTGA	28800
OY	3053	ATGAGAGAGAGAGCGCTTGATCACTAGCATAGGTTTCACACAAGCCCTGGATTTCAGAG	31122
Dd	2881	ATGAGAGAGAGAGCGCTTGATCACTAGCATAGGTTTCACACAAGCCCTGGATTTCAGAG	29400
OY	3113	TGTTTAAACAGAGCTGCGCTCTTTCAGAGACACAGTTTCCATTTCCAAAGAGGACTCACTGA	31722

Db	2941	TTTAAACAGAGCGCTGGCCCTTTTCAGGACCAAGTTCCTCAATTCACAAGAGCGCTACGTGA	3000
Qy	3173	GTCCTCTACTCTCACTGGGGGCCAGAGTAAACGACAAATGCGCTTTTATATATAT	3232
Db	3001	GGTCCCTACTCTCACTGGGGGCCAGAGTAAACGACAAATGCGCTTTTATATATAT	3060
Qy	3233	TTATTTGGTGTGCTCTGTATTATTAAGAGATCAAAATGTATAACGACAGCTCTTTTCAC	3292
Db	3061	TTATTTGGTGTGCGCTGTATTATTAAGAGATCAAAATGTATAACGACAGCTTTTCAC	3120
Qy	3293	CTGACTTGTATACTATACTACTAGTGTGGTGGCTGGGTGTACTTCTCACTGAC	3352
Db	3121	CTGACTTGTATACTATACTACTAGTGTGGTGGGTGTACTTCTCACTGAC	3180
Qy	3353	CGCTAGATAAACGTGTGCGCTGTCCCCCA	3380
Db	3181	CGCTAGATAAACGTGTGCGCTGTCCCCCA	3208

RESULT 8	AK026028	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
	AK026028	2113 bp	linear				
		Homo sapiens CDNA: FLJ22375 fis, clone HRC06785.					
	AK026028	1 GI:10438737					
		Oligo capping; fis (full insert sequence).					
		Homo sapiens primary human renal epithelial cells CDNA to mRNA.					

ORGANISM

REFERENCE
AUTHORS
1 (sites)
Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., . . .

TITLE

AUTHORS
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.

University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,

COMMENT

COMMENT
NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES	Location/Qualifiers
source	1. .2113

BASE COUNT	585 a	538 c	470 g	520 t
ORIGIN	/cell_type="primary human renal epithelial cells" /clone_lib="HRC" /note="cloning vector pME18SFL3"			

Query Match	30.4%	Score 1875.8	DB 9	length 2113
Best Local Similarly	99.8%	Pred. NO. 0		
Matches 1888	Conservative	0	Mismatches 2	Indels 1
				Gaps 1

QY	2448	2507
Db	1	60
	TTAGAAAGGGCGCAAGACAATGACTCCCATGTGTATGCAGTCATCGAGGACACCAGGT	

2508 ATATGGGCATCTGCTACAGGATTCAGCGCTCTTCTCTGACCGACAGAGTGGACACCTA 2567
|||||
61 ATATGGGCATCTGCTACAGGATTCAGCGCTCTTCTCTGACCGACAGAGTGGACACCTA 120
2568 CCGGCGCTTCCAGGGCACCAGTGGGGTCTGTCTCTCTCCACCACCCACCATATGCTCCAG 2627
|||||
121 CCGGCGCTTCCAGGGCACCAGTGGGGTCTGTCTCTCTCCACCACCCACCATATGCTCCAG 180
2628 GGGCCCACTGCAAAATTTGGCCACATGAGAGCCACTCTCGTCCGTCCTCCGACAGTCTGA 2687
|||||
181 GGGCCCACTGCAAAATTTGGCCACATGAGAGCCACTCTCGTCCGTCCTCCGACAGTCTGA 240
2688 GAGTGAACCGTACACCTTCTCCATCCCAACATGGGGATGTAAAGCAGCAAGGACACAGA 2747
|||||
241 GAGTGAACCGTACACCTTCTCCATCCCAACATGGGGATGTAAAGCAGCAAGGACACAGA 300
2748 CATTCCTTACTGACACTCAGAGCCCATGAGACGACAGCAAGAAATTTGATTCATTTCA 2807
|||||
301 CATTCCTTACTGACACTCAGAGCCCATGAGACGACAGCAAGAAATTTGATTCATTTCA 360
2808 GAGCGTTGCTGAGTTCAATTAAGAGGACACTGAGACACCCCTCGTCTCTAACCAAG 2867
|||||
361 GAGCGTTGCTGAGTTCAATTAAGAGGACACTGAGACACCCCTCGTCTCTAACCAAG 420
2868 AATCTCTAAGAGAGGAATTTATACAGAAAGACAGAGAGGTTTCTCTGACACCGCC 2927
|||||
421 AATCTCTAAGAGAGGAATTTATACAGAAAGACAGAGAGGTTTCTCTGACACCGCC 480
2928 AACTTCACATTTGCTCAGTGGACTCATTTAAAGGCAAGACATTTGAAATGTAAATTTCA 2987
|||||
481 AACTTCACATTTGCTCAGTGGACTCATTTAAAGGCAAGACATTTGAAATGTAAATTTCA 540
2988 ACTTGATATACAGTATGACAGCTCATGTGCTCTCAACTTAAAGCTGTGGGGTATAGCCAGC 3047
|||||
541 ACTTGATATACAGTATGACAGCTCATGTGCTCTCAACTTAAAGCTGTGGGGTATAGCCAGC 600
3048 CTGTAAATGAGAGAGAGAGGCTGAGTCACTAGCATAGGGTTGGACAGAGCCCTGGATT 3107
|||||
601 CTGTAAATGAGAGAGAGAGGCTGAGTCACTAGCATAGGGTTGGACAGAGCCCTGGATT 660
3108 CAGAGTGTAAACAGAGGCTTGCCCTTTCAGAGACAAGTGTCAATTTCCAGAGAGACCTA 3167
|||||
661 CAGAGTGTAAACAGAGGCTTGCCCTTTCAGAGACAAGTGTCAATTTCCAGAGAGACCTA 720
3168 CCTGAGAGTCCCTACTCTACTGGGGTCCCGAGAGATGAAACCGACATGTGCTTTTATT 3227
|||||
721 CCTGAGAGTCCCTACTCTACTGGGGTCCCGAGAGATGAAACCGACATGTGCTTTTATT 780
3228 ATTATTTATTTGGTGGTCTGTATTTAAAGATCAATGTATTAACCACTAGGCTTT 3287
|||||
781 ATTATTTATTTGGTGGTCTGTATTTAAAGATCAATGTATTAACCACTAGGCTTT 840
3288 TTCACCTGACTTAAATTAATCACTAATGCTGTTGGATGCTGGTGTGACTTTCTA 3347
|||||
841 TTCACCTGACTTAAATTAATCACTAATGCTGTTGGATGCTGGTGTGACTTTCTA 900
3348 CTGACCGCTAGATTAACGCTGTGCTCTCCCGACAGTGTGGGAATTAATTTACATCTGTC 3407
|||||
901 CTGACCGCTAGATTAACGCTGTGCTCTCCCGACAGTGTGGGAATTAATTTACATCTGTC 960
3408 CAACCAAGAAAGATGTGTGTGTGAGCAGACTGACACATTCCTGCTTTGTAAGAG 3467
|||||
961 CAACCAAGAAAGATGTGTGTGTGAGCAGACTGACACATTCCTGCTTTGTAAGAG 1020
3468 CTTCCTGATTTCTAGAGTGGTTCGTGTATTCCTATTTGGAATTTCAATTTGAAATCCC 3527
|||||
1021 CTTCCTGATTTCTAGAGTGGTTCGTGTATTCCTATTTGGAATTTCAATTTGAAATCCC 1080
3528 ATTGTCTTACTAGTCCCTAGCAATTAAGAAATTTCCCTCAAGTTTCCATGTGCGGTTCTCT 3587
|||||
1081 ATTGTCTTACTAGTCCCTAGCAATTAAGAAATTTCCCTCAAGTTTCCATGTGCGGTTCTCT 1140
3588 ACCTGCAGCAATCTTTGACATTTAAAGAAATTTAGAAATATTCCTACTTAATA 3647

|||||
Db 1141 ACCTGCAGCAATCTTTGACATTTAAAGAAATTTAGAAATATTCCTACTTAATA 1200
3648 ATGTTTAATATATACCAAAACAGTGGCCCCCTGCATATGTTTCTGTGGCACTGCAAC 3707
|||||
1201 ATGTTTAATATATACCAAAACAGTGGCCCCCTGCATATGTTTCTGTGGCACTGCAAC 1260
3708 CATTTACTGGTACTTAAACAAACATATAGTTATAGTCCGCGGGATCAGAAATTCAA 3767
|||||
1261 CATTTACTGGTACTTAAACAAACATATAGTTATAGTCCGCGGGATCAGAAATTCAA 1320
3768 AATGATATGCTCCGTAATGAAATCAAGTGTGACAGAGCTGTGCTCTTCTGAAAGCTC 3827
|||||
1321 AATGATATGCTCCGTAATGAAATCAAGTGTGACAGAGCTGTGCTCTTCTGAAAGCTC 1380
3828 TAGGAGAAAGCCGGTTCCTTCCATTTCAAGCTTTTGAAGCTGTGCTGCAATTTCCAGCT 3887
|||||
1381 TAGGAGAAAGCCGGTTCCTTCCATTTCAAGCTTTTGAAGCTGTGCTGCAATTTCCAGCT 1440
3888 CCAGTGGCTGTCAGCTTTCTACATGATGATGACATGTCAGCTGACCTGCCACTTCC 3947
|||||
1441 CCAGTGGCTGTCAGCTTTCTACATGATGATGACATGTCAGCTGACCTGCCACTTCC 1500
3948 CTCTTTGACTTACAAAGCCACCAGAGAAATCCAGGATTAATCTCTCATTTAAAGATCT 4007
|||||
1501 CTCTTTGACTTACAAAGCCACCAGAGAAATCCAGGATTAATCTCTCATTTAAAGATCT 1560
4008 TCATCATCTCTGAAAGACCTTTTCCATGCAACAAATAGCCACAGAGTGGGGATTAAG 4067
|||||
1561 TCATCATCTCTGAAAGACCTTTTCCATGCAACAAATAGCCACAGAGTGGGGATTAAG 1620
4068 ACCAGACATCTTTGGGGTCTGTTATTTGCTTACACACCTTCTGCCACATGACTCCC 4127
|||||
1621 ACCAGACATCTTTGGGGTCTGTTATTTGCTTACACACCTTCTGCCACATGACTCCC 1680
4128 ACAGAGAGGCTCAAAATATATCTGGGACAGAGGATGTTTGAAGCTTGGGACTCT 4187
|||||
1681 ACAGAGAGGCTCAAAATATATCTGGGACAGAGGATGTTTGAAGCTTGGGACTCT 1740
4188 AACACTT-AAAAAACCCGATCAGAAATGTCGGCCATGCTGGGGCTCAATTTCCACC 4246
|||||
1741 AACACTTAAAAAAACCCGATCAGAAATGTCGGCCATGCTGGGGCTCAATTTCCACC 1800
4247 TAGCAACAATGGCTGAGCTGGGACACACCTGCTTTGAAGAGGGTGTCCACTTCC 4306
|||||
1801 TAGCAACAATGGCTGAGCTGGGACACACCTGCTTTGAAGAGGGTGTCCACTTCC 1860
4307 CAGTCAACCAAGCCCACTACGCTATC 4337
|||||
1861 CAGTCAACCAAGCCCACTACGCTATC 1891
Db

RESULT 9
AK026187
LOCUS AK026187 1241 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22534 fls. clone HNCI3020.
ACCESSION AK026187
VERSION AK026187.1 GI:10438956
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens primary human renal epithelial cells cDNA to mRNA,
clone J1b-HRC clone:HRC13020.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hiki,I., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEBO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1241)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- 8 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES
source Location/Qualifiers
1..1241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC13020"
/cell_type="primary human renal epithelial cells"
/clone_1lb="HRC"
/notes="cloning vector pME18Sfl3"
65..>1241
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15388.1"
/db_xref="gi:10438957"
/translation="MAGNFNLISGCGDQDASPILRLQFVLVQHPQNESNKIYVDDLSNRSASLITIEPRPVKOSRKFVSGFVCLSPSCSSMLTSSSKRISPLCODLRLRMVAHEKTIESTDHRGCOORSYSLQVSDIHLHLEVLHDFEMKILVPRDLSTLVPARORLOOHEKRCMPTSEFVYASAIKPSODIYSGSCBPSGSIQIOYKONISVLTLPFASFOQDASRQGLITVFIPFKKEGVFTYPTDKSVYLLTFPMWDGLSLTSSVNNISVPRDDCTFFPKERSGVCCGTFGRAFMIIQERTAEELFSLDEVLLKPSFHHHSFWVYNISCSPTSGKQDLDFSVLTLPRTVLTLLIAVGGVLLSALGLITCVKKRRK"

QY	1613	ATGACTTTCCTCGAAGCGCTGCGGCCAAGGACGAGCTCAGCTGGTGGTGGCCAG	1672
Db	483	ATGACTTTCCTCGAAGCTGCTGGTGGCCAGAGACAGGCTCAGCTGGTGGTGGCCAG	542
QY	1673	CCCAAGAGCTGCAGAGAGATACACAGAAAGCCCTGCACACAGCTTCAGCTACTCG	1732
Db	543	CCCAAGAGGCTGCAGAGAGATACACAGAAAGCCCTGCACACAGCTTCAGCTACTCG	602
QY	1733	TGGCCAGTGCATACCAGACCAGAGACTGTACTTGGGCTCCTTCGCCCCGAGGCTCTA	1792
Db	603	TGGCCAGTGCATACCAGACCAGAGACTGTACTTGGGCTCCTTCGCCCCGAGGCTCTA	662
QY	1793	TCAACGAGTCCAGGTGGAAGCAGAACATCTGGGTGATACCCCTTCGACCTTTGGCCCCAGCT	1852
Db	663	TCAACGAGTCCAGGTGGAAGCAGAACATCTGGGTGATACCCCTTCGACCTTTGGCCCCAGCT	722
QY	1853	TCCAACAGAGAGCTCCACAGGAGGCTGTGACGGTCTCCTTTATACCTTATTTCAAGAAG	1912
Db	723	TCCAACAGAGAGCTCCACAGGAGGCTGTGACGGTCTCCTTTATACCTTATTTCAAGAAG	782
QY	1913	AAGCGCTTTTCAGGCTGACCCCTGACACAAAAAGCAGGCTCTACCTGAGAGACCCCAACT	1972
Db	783	AAGCGCTTTTCAGGCTGACCCCTGACACAAAAAGCAGGCTCTACCTGAGAGACCCCAACT	842
QY	1973	GGGAGCCGGGGGCTGTCATACCTCCTGACCTGTGTCTGTGAAACATCAAGCGTCCAGAGACC	2032
Db	843	GGGAGCCGGGGGCTGTCATACCTCCTGACCTGTGTCTGTGAAACATCAAGCGTCCAGAGACC	902
QY	2033	AGGTGGCTGCTGACTTCTTTAAGAGAGCGAGCGGCGTGTGTCGCCAGACAGGCGCG	2092
Db	903	AGGTGGCTGCTGACTTCTTTAAGAGAGCGAGCGGCGTGTGTCGCCAGACAGGCGCG	962
QY	2093	CATTATGATCATCCAGAGACAGCGGACCCGGGCTGAGAGATCTTCAGCTGACGAGG	2152
Db	963	CATTATGATCATCCAGAGACAGCGGACCCGGGCTGAGAGATCTTCAGAGCTGACGAGG	1022
QY	2153	ATGTGCTCCCCCAAGGCCAAGCTTCACACATCAGAGTTCTGGGTCAACAATCTCTACTGCA	2212
Db	1023	ATGTGCTCCCCCAAGGCCAAGCTTCACACATCAGAGTTCTGGGTCAACAATCTCTACTGCA	1082
QY	2213	GCCCCAGCAGCGCGCAGCAGAGCTAGACCTGCTCTTCGCGGTGACACTTACCCCAAGAGCTG	2272
Db	1083	GCCCCAGCAGCGCGCAGCAGAGCTAGACCTGCTCTTCGCGGTGACACTTACCCCAAGAGCTG	1142
QY	2273	TGGACTTGACTGTCAATCCTCATCGCAGCGGTGGAGGTGGAGTCTTACTGCTGTGCC	2332
Db	1143	TGGACTTGACTGTCAATCCTCATCGCAGCGGTGGAGGTGGAGTCTTACTGCTGTGCC	1202
QY	2333	TGGGCGTCATCATTTGCTGTGTGAAGAAAGAAAGAAAG 2371	
Db	1203	TGGGCGTCATCATTTGCTGTGTGAAGAAAGAAAGAAAG 1241	
RESULT 10			
BC021099			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
<p>BC021099 1410 bp mRNA linear PRI 22-JAN-2002</p> <p>Homo sapiens, Similar to hypothetical protein FLJ22969, clone</p> <p>Submitted (03-JAN-2002), Cancer Genomics Office, National Cancer</p> <p>Gene Collection (MGC), Cancer Genomics Office, Health, Mammalian</p> <p>Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,</p> <p>USA</p> <p>NIH-MGC Project URL: http://mgc.ncl.nih.gov</p>			

COMMENT

Contact: MGC help desk
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: DCTD/DPF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@gscc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Lang, Carlie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 40 Row: f Column: 14
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 12383093.
 Location/Qualifiers

FEATURES

source

1..1410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:31813 IMAGE:4590554"
 /tissue_type="Kidney, renal cell adenocarcinoma"
 /clone_1lb="NH_MGC_14"
 /lab_host="DH10B-R"
 /note="Vector: pOT87"
 136..1167
 /codon_start=-1
 /product="Similar to hypothetical protein FLJ22969"
 /protein_id="AAH21099.1"
 /db_xref="gi:18088979"
 /translation="MAGLNGVSIALLGLVLLGAARLPGRGAFETALPRENSITVLI
 KICPTLAKPCYIVISKRHTMISKRSERIVFTFSCSPENHRIYETOKNIDCSG
 PCPEGVQDPTSLPTLNRTFIMVNAHKSIGLOPSIFRLRIGIGESCDDVT
 HSIQRIDATVVRIGTFCNSGTVSRIKMEGVKMLHLFWPHRNVSGFINRSRISIK
 RLCTIESVEEGGSATLMSANYPEGPEDELTWQFVVAHRLVSVSEFINLSNCR
 KEEREVYIPGSTNPVEVFKLBDKQGNMAGNPNLSLQCDQDAGSPGILRLQFQVLV
 QHPNNESS"

CDS

BASE COUNT

321 a 395 c 370 g 324 t

ORIGIN

Query Match 18.8%; Score 1159.8; DB 9; Length 1410;
 Best Local Similarity 99.8%; Pred. No. 1.8e-233;
 Matches 1161; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 148 GGGCGGGGCTCGGGCGGCTCCGCGCGGCGAGGTGAGTGAAGCGGCGGAGCGCAGCT 207
 |||||||
 DB 1 GGGCGGGGCTCGGGCGGCTCCGCGCGGCGAGGTGAGTGAAGCGGCGGAGCGCAGCT 60
 QY 208 GGGCGGGGCTTGGGGCGCTGGGGCGGCGGCTCCGCGCGGCTGTTTCCCGACGAGGCGG 267
 |||||||
 DB 61 GGGCGGGGCTTGGGGCGCTGGGGCGGCGGCTCCGCGCGGCTGTTTCCCGACGAGGCGG 120
 QY 268 AGCGCTCCCGGAGTCAATGCGCGGCTGAATCGGGGGTCTCTATCGCACTGCTAGGGGTT 327
 |||||||
 DB 121 AGCGCTCCCGGAGTCAATGCGCGGCTGAATCGGGGGTCTCTATCGCACTGCTAGGGGTT 180
 QY 328 CTGCTCTGCGGTGGCGGCGGCTGCGCGCGGCGGAGAGCTTTTGAGATGCTCTGCGCA 387
 |||||||
 DB 181 CTGCTCTGCGGTGGCGGCGGCTGCGCGCGGCGGAGAGCTTTTGAGATGCTCTGCGCA 240
 QY 388 CGAGAAAGCAACATTCAGTTCTCATAAAGCTGGGAGCCCGCGACTGCTGGGAAAAACC 447
 |||||||
 DB 241 CGAGAAAGCAACATTCAGTTCTCATAAAGCTGGGAGCCCGCGACTGCTGGGAAAAACC 300
 QY 448 TGTATACATGCTCATTTCTAAAGACATATACCATGTTGTCTCATCAAGTCTGGAGAAGA 507
 |||||||
 DB 301 TGTATACATGCTCATTTCTAAAGACATATACCATGTTGTCTCATCAAGTCTGGAGAAGA 360

QY 508 ATAGCTTTACCTTTAGCTTCCAGAGTCGAGATGCTTGTGTCATAGAGATCCAGAAA 567
 |||||||
 DB 361 ATAGCTTTACCTTTAGCTTCCAGAGTCGAGATGCTTGTGTCATAGAGATCCAGAAA 420
 QY 568 AATATTGACATGTAATGTCAGAGCCCATGCTCTTTGGGAGGTTGACGTTACGCCCTGACA 627
 |||||||
 DB 421 AATATTGACATGTAATGTCAGAGCCCATGCTCTTTGGGAGGTTGACGTTACGCCCTGACA 480
 QY 628 TCGTTTGGCTTACCTTCAACAGAACTTTCATCTGGGATGTCAAAGCTCATTAAGAGCATC 687
 |||||||
 DB 481 TCGTTTGGCTTACCTTCAACAGAACTTTCATCTGGGATGTCAAAGCTCATTAAGAGCATC 540
 QY 688 GGTTAGAGTCGAGTTTTCATTCCTCGGCTGAGGAGATCGGTCGGGTTGAGAGCTGC 747
 |||||||
 DB 541 GGTTAGAGTCGAGTTTTCATTCCTCGGCTGAGGAGATCGGTCGGGTTGAGAGCTGC 600
 QY 748 CCAAGAGGATCACTCACTCATCAGCGGCGGATGATGCCACCGGTGAGAGATCGGA 807
 |||||||
 DB 601 CCAAGAGGATCACTCACTCATCAGCGGCGGATGATGCCACCGGTGAGAGATCGGA 660
 QY 808 ACCTTTGCAGCAATGCGCACTGTGTCGCGGATCAAGATGCAAGAGAGTGAAGATGGCC 867
 |||||||
 DB 661 ACCTTTGCAGCAATGCGCACTGTGTCGCGGATCAAGATGCAAGAGAGTGAAGATGGCC 720
 QY 868 TTACACCTCCCATGTTTCCACCCCAAAATGCTCCGCGTTCCAGATGCAAAACCGCTCA 927
 |||||||
 DB 721 TTACACCTCCCATGTTTCCACCCCAAAATGCTCCGCGTTCCAGATGCAAAACCGCTCA 780
 QY 928 TCTATATAAGAGTCTGTGTCATCATGATGATGTTGTGAGGAGTGAAGGCTGAGCAACCTG 987
 |||||||
 DB 781 TCTATATAAGAGTCTGTGTCATCATGATGATGTTGTGAGGAGTGAAGGCTGAGCAACCTG 840
 QY 988 ATGTCTGCCACATACCCAGAGGCTTCCAGAGATGAGCTCATGAGTGGAGTTTGTG 1047
 |||||||
 DB 841 ATGTCTGCCACATACCCAGAGGCTTCCAGAGATGAGCTCATGAGTGGAGTTTGTG 900
 QY 1048 GTTCTCTGACACCTGCGGGGCGGAGCTCTCTCTCAACTTCAACCTCTCCAACTGTGAG 1107
 |||||||
 DB 901 GTTCTCTGACACCTGCGGGGCGGAGCTCTCTCTCAACTTCAACCTCTCCAACTGTGAG 960
 QY 1108 AGAAGAGAGAGGAGGTTGAATCACTATCCCGGGCTCCACCAACCAACCCGAGAGTTTC 1167
 |||||||
 DB 961 AGAAGAGAGAGGAGGTTGAATCACTATCCCGGGCTCCACCAACCAACCCGAGAGTTTC 1020
 QY 1168 AAGCTGAGAGACAAGACGCTTGGGAACATGAGCGGGAATCTCAACTCTCTGCAAGGC 1227
 |||||||
 DB 1021 AAGCTGAGAGACAAGACGCTTGGGAACATGAGCGGGAATCTCAACTCTCTGCAAGGC 1080
 QY 1228 TGTGACCAAGATGCCCAAAATGCCAGGATCTCGGCTGCAGTTCCAAAGTTTGTGTCAA 1287
 |||||||
 DB 1081 TGTGACCAAGATGCCCAAAATGCCAGGATCTCGGCTGCAGTTCCAAAGTTTGTGTCAA 1140
 QY 1288 CATCCACAAATGAAGCAATTA 1310
 |||||||
 DB 1141 CATCCACAAATGAAGCAATGA 1163

RESULT 11

AX464028

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Sequence 161 from Patent WO0140466.
 AX464028
 AX464028.1 GI:21899031
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1
 Baker, K.P., Beresini, M., DeForge, L., Desnoyers, L., Filvaroff, E.,
 Gao, W.O., Gerlitsen, M.E., Goddard, A., Godowski, P.J., Gurney, R.L.,
 Sherwood, S., Smith, V., Stewart, T.A., Tumas, D., Watanabe, C.K.,

Wood, W.L. and Zhang, Z.
 Secreted and transmembrane polypeptides and nucleic acids encoding
 same
 JOURNAL Patent: WO 0140466-A 161 07-JUN-2001;
 Genentech Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 311 a 384 c 352 g 329 t
 ORIGIN

Query Match 18.1%; Score 1117.8; DB 6; Length 1376;
 Best Local Similarity 99.8%; Pred. No. 1.3e-224;
 Matches 1119; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

190 CCAGGGGAGAGCAGCTGCGCGGCTTGGGCGCTGGGGCGCGCGCTCCCAACCGTGC 249
 1 CCAGGGGAGAGCAGCTGCGCGGCTTGGGCGCTGGGGCGCGCGCTCCCAACCGTGC 60

250 TTTTCCCGACCGAGCGGCGGCTCCCGAGTCATGGCGGCTGAACCTGGGGGTCTCT 309
 61 TTTTCCCGACCGAGCGGCGGCTCCCGAGTCATGGCGGCTGAACCTGGGGGTCTCT 120

310 ATGCGACTGCTAGGGGTTCTCTGCTGGGTGGCGGCGCGCTCCCGCGGGGAGAGCT 369
 121 ATGCGACTGCTAGGGGTTCTCTGCTGGGTGGCGGCGCGCTCCCGCGGGGAGAGCT 180

370 TTTAGAGTTGCTGTCGACGAGAAACATTTACAGTTCTCATTAAGCTGGGAGCCCG 429
 181 TTTAGAGTTGCTGTCGACGAGAAACATTTACAGTTCTCATTAAGCTGGGAGCCCG 240

430 ACTCTGCTGGCAAAACCTGTTCATCGTCAATTTCTAAAGACATATAACCATGTTC 489
 241 ACTCTGCTGGCAAAACCTGTTCATCGTCAATTTCTAAAGACATATAACCATGTTC 300

490 ATCAAGCTGGAGAAATATAGTTTACCTTTAGCTGGCGAGTCCCTGAGAACTCATT 549
 301 ATCAAGCTGGAGAAATATAGTTTACCTTTAGCTGGCGAGTCCCTGAGAACTCATT 360

550 GTCATAGAGATCCAGAAATATATGACTGTATGTCAGAGCCCATGCTTTGGGGAGGT 609
 361 GTCATAGAGATCCAGAAATATATGACTGTATGTCAGAGCCCATGCTTTGGGGAGGT 420

610 CAGCTTACGCGCTGACATGTTGTTGCTCAACCTCAACAGAACTTTCATCTGGAGTTC 669
 421 CAGCTTACGCGCTGACATGTTGTTGCTCAACCTCAACAGAACTTTCATCTGGAGTTC 480

670 AAAGCTATAGAGATCCAGAAATATATGACTGTATGTCAGAGCCCATGCTTTGGGGAGGT 729
 481 AAAGCTATAGAGATCCAGAAATATATGACTGTATGTCAGAGCCCATGCTTTGGGGAGGT 540

730 GGTCCGGGTGAGAGCTGCCAGAGAGATCACTCATCTCCATCAGCGGCGGAATGATGCC 789
 541 GGTCCGGGTGAGAGCTGCCAGAGAGATCACTCATCTCCATCAGCGGCGGAATGATGCC 600

790 ACCGTGTCAAGATCGAATCCTTCTGACGAATGGCACTGTGTCCGGATCAAGATCAA 849
 601 ACCGTGTCAAGATCGAATCCTTCTGACGAATGGCACTGTGTCCGGATCAAGATCAA 660

850 GAAGAGATGAAATAGGCGCTTACACCTCCCATGTTCCACCCCAAAATGTCTCCGGCTTC 909
 661 GAAGAGATGAAATAGGCGCTTACACCTCCCATGTTCCACCCCAAAATGTCTCCGGCTTC 720

910 AGCATTCGAAACCGCTCATCTATAAAGCTGTGTCATCATGATCGATGTTGGAGGT 969
 721 AGCATTCGAAACCGCTCATCTATAAAGCTGTGTCATCATGATCGATGTTGGAGGT 780

970 GAAGGCTCAGCAACCTGATGTCTGCCAATCCAGAAAGGCTTCCCTGAGAGATGAGCTC 1029
 781 GAAGGCTCAGCAACCTGATGTCTGCCAATCCAGAAAGGCTTCCCTGAGAGATGAGCTC 840

1030 ATGAGCTGGAGTTGTGCTTCTGTCACACCTGGGGGCGAGCTCTCTTCTCAACTTC 1089

|||||
 Db 841 ATGAGCTGGGAGTTGTCTCTGTCACACCTGGGGGCGAGCTCTCTTCTCAACTTC 900

1090 AACCTTCGAACTGAGAGAGAGAGGAGGAGGTTGAATCTCATCTCCGGGCTCCACC 1149
 Db 901 AACCTTCGAACTGAGAGAGAGAGGAGGAGGTTGAATCTCATCTCCGGGCTCCACC 960

1150 ACCAACCCTGAGGTTTCAGCTGAGAGACAGAGCTGGGAACATGCGGGGAATTC 1209
 Db 961 ACCAACCCTGAGGTTTCAGCTGAGAGACAGAGCTGGGAACATGCGGGGAATTC 1020

1210 AACCTTCCTGCAAGGCTGTGACCAAGATGCCCAAAAGTCCAGGATCTCCGGCTGAG 1269
 Db 1021 AACCTTCCTGCAAGGCTGTGACCAAGATGCCCAAAAGTCCAGGATCTCCGGCTGAG 1080

1270 TTCCAGTTTGTGTCACATCCACAAATGAAGCAATTA 1310
 Db 1081 TTCCAGTTTGTGTCACATCCACAAATGAAGCACTGA 1121

RESULT 12
 AK026329 976 bp mRNA linear PRI 29-SEP-2000
 LOCUS AK026329
 DEFINITION Homo sapiens cDNA: FLJ22676 fis, clone HS110569.
 ACCESSION AK026329
 VERSION AK026329.1 GI:10439158
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens human small intestine cDNA to mRNA, clone_11b.HSI
 clone:HS110569.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)

REFERENCE

Matanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
 Ota, T., Suzuki, Y., Oba-yashi, M., Nishi, T., Shihabara, T., Tanaka, T.,
 Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 976)

JOURNAL

Shihabara, T., Tanaka, T. and Nakamura, Y.
 Direct Submission
 Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure Analysis, Human
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
 Japan (E-mail:cdna@lelms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan. cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing: Departent of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES

source 1..976
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HS110569"
 /tissue_type="human small intestine"
 /clone_11b="HS1"
 /note="cloning vector pME18SFL3"
 BASE COUNT 251 a 257 c 219 g 249 t
 ORIGIN

Query Match 15.6%; Score 959; DB 9; Length 976;
 Best Local Similarity 99.5%; Pred. No. 3.3e-191;
 Matches 962; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

4956 AATGTAATCATTTGATTTCAATGATTAATCACTTGAATAGCAGCGCTCAACTGTTCA 5015
 1 AATGTAATCATTTGATTTCAATGATTAATCACTTGAATAGCAGCGCTCAACTGTTCA 60

```

OY 5016 CATGGCAATGAAGAGATTTTATATGTCCTCTAAATGGCGTGTACTGCAGACCTCTT 5075
DB 61 CATGGCATTAATGAAGAGATTTTATATGTCCTCTAAATGGCGTGTACTGCAGACCTCTT 120
OY 5076 GAACACTTTCCAGAGAGATGATATTTAATGATCCCTTGGCTTGGCTTATGGACCTT 5135
DB 121 GAACACTTTCCAGAGAGATGATATTTAATGATCCCTTGGCTTGGCTTATGGACCTT 180
OY 5136 TCCCTTTCTAAGAGTGTGTTCTTCCAGTACCTTGGCTTGTGAGCCGAGATCTGA 5195
DB 181 TCCCTTTCTAAGAGTGTGTTCTTCCAGTACCTTGGCTTGTGAGCCGAGATCTGA 240
OY 5196 CCTGTGATTAAGGGCCAAAGAGAGGCGTGGGCTTCTCTCCCTACGAAAGCCCTATT 5255
DB 241 CCTGTGATTAAGGGCCAAAGAGAGGCGTGGGCTTCTCTCCCTACGAAAGCCCTATT 300
OY 5256 TGAATTTCACTGTGTGAGGCGCTAGCCCTCATTTCTGACATTTCCCAACCTCCACGCC 5315
DB 301 TGAATTTCACTGTGTGAGGCGCTAGCCCTCATTTCTGACATTTCCCAACCTCCACGCC 360
OY 5316 TTCCAAAGCAGAGTAGTGTGCTTCCGATTCACCCAAAGTGGGATTTGGCTTCTTACGCT 5375
DB 361 TTCCAAAGCAGAGTAGTGTGCTTCCGATTCACCCAAAGTGGGATTTGGCTTCTTACGCT 420
OY 5376 GGGTACTTGTCAACCATCAGCATGCTGTTGGCTTCCGAAAGGAGACCGTGGCATTTT 5435
DB 421 GGGTACTTGTCAACCATCAGCATGCTGTTGGCTTCCGAAAGGAGACCGTGGCATTTT 480
OY 5436 CCTTCACTGAGGGCTCAAAAACCTCTGAGACAGTTGCTGCTTCTGAGACAGATATTTCC 5495
DB 481 CCTTCACTGAGGGCTCAAAAACCTCTGAGACAGTTGCTGCTTCTGAGACAGATATTTCC 540
OY 5496 TGGAGCTGTGCTCACTGATGAAGGGGCCAGCTGTGAGAACCCCTGCTTTCTTAAAGC 5555
DB 541 TGGAGCTGTGCTCACTGATGAAGGGGCCAGCTGTGAGAACCCCTGCTTTCTTAAAGC 600
OY 5556 CCAGGCGCCCACTTACATTAACATTTCAAGGCTCACTGGAACAGTGAAGGCCATTTGTT 5615
DB 601 CCAGGCGCCCACTTACATTAACATTTCAAGGCTCACTGGAACAGTGAAGGCCATTTGTT 660
OY 5616 GAAGCTTACTGATGCGCAGCCCACTGCTATCCAGCTGTGCTTCCATTCAGAGGAAG 5675
DB 661 GAAGCTTACTGATGCGCAGCCCACTGCTATCCAGCTGTGCTTCCATTCAGAGGAAG 720
OY 5676 GCCAGGCGATGCAAGAGCTGCTCTAATGCTGTGCTCATTTGCCACAGAGGAAGCTCTC 5735
DB 721 GCCAGGCGATGCAAGAGCTGCTCTAATGCTGTGCTCATTTGCCACAGAGGAAGCTCTC 780
OY 5736 AAGGAAGAGTCAACTGGGACAGACAGCAAGCCGAGCATGGCTTGGTAAAGGTTAGC 5795
DB 781 AAGGAAGAGTCAACTGGGACAGACAGCAAGCCGAGCATGGCTTGGTAAAGGTTAGC 840
OY 5796 AGACTGTGTGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5855
DB 841 AGACTGTGTGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 900
OY 5856 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5915
DB 901 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960
OY 5916 GTATCA 5922
DB 961 AAAAAA 967

```

```

RESULT 13
LOCUS BC011340 3843 bp mRNA linear ROD 07-AUG-2002
DEFINITION Mus musculus, clone IMAGE:3709937, mRNA, partial cds.
ACCESSION BC011340
VERSION BC011340.1 GI:15030169
KEYWORDS
SOURCE house mouse.

```

```

ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 3843)
JOURNAL Strausberg, R.
Direct Submission
Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cghabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAC Plate: 17 Row: 1 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
source location/Qualifiers
1..3843
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3709937"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
<1..1155
/codon_start=1
/product="Unknown (protein for IMAGE:3709937)."
/protein_id="AAH1340.1"
/db_xref="GI:15030170"
/translation="LLVPKDKLSLMLVPGOKLQHTOERPCNTSFGHYVSTTPGDL
YRSPGSGSISIKIOYKONSSVTLRAYAPSPQEVSKGLIYSTPYREGEIFTPTP
DPRNKVYLSPMDRDLPLSLSSVSNISPSNOVACTYLKRSGLACSGAPMIIQ
EQQSRBELFSLSEEVLPKPSFHHSFVNISCSPMNKQDLLEWVLTLPRTYDLA
VVIAGAGGALLFLVLLIFVKKKRKYDKPAGVINGVNTQMPPQPKPPKRD
NDSHYAVIEDPMVYGHLLDQSGSGLQPEVDTYRFFQSPMDGCPPTPLPSRPTA
KPTAEIAPLSSPSESEBPTFESHPKSGEIGVRENDIPLLHFGVETEE"
BASE COUNT 1015 a 943 c 906 g 979 t
ORIGIN
Query Match 13.3%; Score 821.2; DB 10; Length 3843;
Best Local Similarity 74.1%; Pred. No. 3e-162;
Matches 1183; Conservative 0; Mismatches 378; Indels 35; Gaps 10;
OY 1630 CTGCTGGTCCCAAGAGAGCTCAGCTGTGCTGTGCTGACCCAGAGAGCTGACGAC 1689
DB 1 CTACTGGTCCCAAGAGAGCTCAGCTGTGCTGTGCTGACCCAGAGAGCTGACGAC 60
OY 1690 CATACACAGAGAGAGCCCTCAACACAGCTTACAGTCTGCGCAGTGCATACCC 1749
DB 61 CATACACAGAGAGAGCCCTTACACACAGCTTGGGACACAGTGCATACACACCC 120
OY 1750 AGCCAGAGCTGTACTTGGCTCTTCCGCGGAGAGCTGTATCAAGAGATCAAGGTG 1809
DB 121 GCCAGAGAGCTGTACTTGGCTCTTCCGCTCAGAGAGGCTCATCGAAGATCCAGGTG 180

```

QY 1810 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1869
 DB 1811 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 240
 DB 181 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 240
 QY 1870 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1929
 DB 241 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 300
 QY 1930 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1989
 DB 301 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 360
 QY 1990 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2049
 DB 361 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 420
 QY 2050 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2109
 DB 421 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 480
 QY 2110 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2169
 DB 481 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 540
 QY 2170 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2229
 DB 541 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 600
 QY 2230 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2289
 DB 601 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2657
 QY 2290 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2349
 DB 658 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 717
 QY 2350 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2409
 DB 718 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 774
 QY 2410 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2469
 DB 775 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 831
 QY 2470 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2529
 DB 832 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 891
 QY 2530 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2589
 DB 892 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 951
 QY 2590 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2649
 DB 952 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1011
 QY 2650 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2709
 DB 1012 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1071
 QY 2710 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2769
 DB 1072 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1131
 QY 2770 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2828
 DB 1132 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1191
 QY 2829 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2888
 DB 1192 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1250
 QY 2889 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 2948

DB 1251 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1307
 QY 2949 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 3008
 DB 1308 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1359
 QY 3009 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 3068
 DB 1360 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1417
 QY 3069 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 3128
 DB 1418 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1466
 QY 3129 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 3188
 DB 1467 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1526
 QY 3189 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 3224
 DB 1527 AAGCAGAACATCTGGGTGACCCCTTGCACCTTTGCCCCAGCTTCCACAAAGAGCCCTCC 1562

RESULT 14
 BC034137
 LOCUS
 DEFINITION
 MGC:32174 IMAGE:5004246, mRNA, complete cds.
 ACCESSION
 BC034137
 VERSION
 BC034137.1 GI:21707874
 KEYWORDS
 MGC.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathia; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 1999)
 Strausberg, R.
 Direct Submission
 Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

REMARK
 COMMENT
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) md@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
 Series: IRAC Plate: 44 Row: P Column: 21
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
 location/Qualifiers

FEATURES
 source
 1..1999
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N-3"
 /clone="MGC:32174 IMAGE:5004246"
 /tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month old mouse, taken by biopsy."
 /clone_1lb="MCI.CGAP_Mam2"
 /lab_host="DH10B"
 /note="vector: pCMV-Sport6"

[illegible]

Db	784	G	784
<hr/>			
RESULT_15	AX192795	544 bp	DNA
LOCUS	AX192795		Linear
DEFINITION	Sequence 362 from Patent WO0149716.		PAT 15-AUG-2001
ACCESSION	AX192795		
VERSION	AX192795.1	GI:15210751	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 544)		
AUTHORS	Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,		
TITLE	Stolk,J.A., King,G.E., Wang,T. and Jlang,Y.		
JOURNAL	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use		
	Patent: WO 0149716-A 362 12-JUL-2001;		
FEATURES	CORIXA CORPORATION (US)		
source	location/Qualifiers		
	1..544		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	138 a 123 c 113 g 170 t		
ORIGIN			
Query Match	8.8%; Score 542.4; DB 6; Length 544;		
Best Local Similarity	99.8%; Pred. No. 1.5e-103;		
Matches 543; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
OY	3068 CCTGAGTACCCTGATGAGGTGGAGACAAGCCCTGCGATTGAGATGTTAAACAGAGCT	3127	
Db	1 CCTGAGTACCCTGATGAGGTGGAGACAAGCCCTGCGATTGAGATGTTAAACAGAGCT	60	
OY	3128 TGCCCTTCACAGAACAGATTCGAATTCAGAAGACCTACTGAGTGCCCTACTCTCAC	3187	
Db	61 TGCCCTTCACAGAACAGATTCGAATTCAGAAGACCTACTGAGTGCCCTACTCTCAC	120	
OY	3188 TGGGGTCCCCAGAGTAGAAAACGACAAATGTCCTTTTAATTATTATTATTGGTGTCCT	3247	
Db	121 TGGGGTCCCCAGAGTAGAAAACGACAAATGTCCTTTTAATTATTATTATTGGTGTCCT	180	
OY	3248 GTGTTATTAAAGATCAAAATGATAACCACTGCTCTTTCACCTGACTAGTAATGA	3307	
Db	181 GTGTTATTAAAGATCAAAATGATAACCACTGCTCTTTCACCTGACTAGTAATGA	240	
OY	3308 CTCATCTAACGATGTTGGATGCCGTGGGTGTGACTTACTGACCGGTAGATAAAGCTG	3367	
Db	241 CTCATCTAACGATGTTGGATGCCGTGGGTGTGACTTACTGACCGGTAGATAAAGCTG	300	
OY	3368 TGCCCTGTCCCCAGAGTGGTGGGAATTAATTAACAATCTGCCAACCGAAAAAGAAATGTGTG	3427	
Db	301 TGCCCTGTCCCCAGAGTGGTGGGAATTAATTAACAATCTGCCAACCGAAAAAGAAATGTGTG	360	
OY	3428 TGTTTGACGAGCATGACACAAATGCTGTTGATTAAGAGACTTCTGTAATCCTAGGTGC	3487	
Db	361 TGTTTGACGAGCATGACACAAATGCTGTTGATTAAGAGACTTCTGTAATCCTAGGTGC	420	
OY	3488 GTTCTGTGTTTCCCATTTGTGGAATTCATCTTGAATCCCATTTGTCATATAGTCTCTACA	3547	
Db	421 GTTCTGTGTTTCCCATTTGTGGAATTCATCTTGAATCCCATTTGTCATATAGTCTCTACA	480	
OY	3548 ATAGAGAAATTTCTCTCAAGTTTCCATGTGCGGTTCTCCTAGCTCAGACATACTTTGAC	3607	
Db	481 ATAGAGAAATTTCTCTCAAGTTTCCATGTGCGGTTCTCCTAGCTCAGACATACTTTGAC	540	
OY	3608 ATTTT 3611		
Db	541 ATTTT 544		

Search completed: November 11, 2002, 09:32:45
Job time : 16209 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:31:42 ; Search time 22 Seconds

(without alignments)
1118.071 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394
Sequence: 1 MAGLNGCSIALGLVLLGA.....SSKDDIDPLNTQEPMEPAE 836

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	2.7	2476	2	US-08-276-967-2
2	116	2.6	1589	3	US-08-755-587-189
3	110.5	2.5	1719	2	US-08-459-568-4
4	110.5	2.5	1719	2	US-08-399-411-4
5	110.5	2.5	1719	3	US-08-516-859A-4
6	110.5	2.5	1719	4	US-09-586-472-4
7	110.5	2.5	1719	4	US-09-528-706-4
8	106.5	2.4	2343	4	US-09-324-867-2
9	106	2.4	670	2	US-08-473-750-11
10	106	2.4	670	2	US-08-477-325-11
11	105.5	2.4	683	4	US-09-213-295D-1
12	104	2.4	449	2	US-08-839-008-2
13	103.5	2.4	727	5	US-08-475-844-9
14	103.5	2.4	727	5	PCR-US95-08429-9
15	103	2.3	449	2	US-08-839-008-9
16	102	2.3	2183	3	US-08-746-111-5
17	101.5	2.3	535	4	US-09-117-860-24
18	100.5	2.3	468	4	US-08-839-008-7
19	100.5	2.3	468	4	US-09-032-523-8
20	98.5	2.2	567	1	US-08-007-2828-2
21	98.5	2.2	1481	2	US-08-616-844-40
22	98.5	2.2	1481	2	US-08-599-654-40
23	98.5	2.2	1481	3	US-08-944-868A-40
24	98.5	2.2	1481	3	US-08-944-423A-40
25	98.5	2.2	1481	3	US-08-944-496-40
26	97.5	2.2	318	4	US-09-068-051A-32
27	97.5	2.2	1785	4	US-09-341-587-3

28	96.5	2.2	1332	2	US-08-971-244-2	Sequence 2, Appl1
29	96.5	2.2	1332	2	US-09-286-891-2	Sequence 2, Appl1
30	96	2.2	401	2	US-08-839-008-5	Sequence 5, Appl1
31	95.5	2.2	424	1	US-08-453-472-8	Sequence 8, Appl1
32	95.5	2.2	424	1	US-08-038-948-2	Sequence 2, Appl1
33	95.5	2.2	424	1	US-08-038-948-5	Sequence 5, Appl1
34	95.5	2.2	424	1	US-08-453-952-8	Sequence 8, Appl1
35	95.5	2.2	424	2	US-08-862-903-8	Sequence 8, Appl1
36	95.5	2.2	703	4	US-09-433-043B-124	Sequence 124, App
37	95.5	2.2	1706	2	US-08-459-568-2	Sequence 2, Appl1
38	95.5	2.2	1706	3	US-08-399-411-2	Sequence 2, Appl1
39	95.5	2.2	1706	3	US-08-516-859A-2	Sequence 2, Appl1
40	95.5	2.2	1706	4	US-09-586-472-2	Sequence 2, Appl1
41	95.5	2.2	1706	4	US-09-528-706-2	Sequence 2, Appl1
42	95	2.2	749	4	US-09-562-737-93	Sequence 93, Appl1
43	95	2.2	4302	3	US-08-658-136-5	Sequence 5, Appl1
44	95	2.2	4302	4	US-09-052-469-8	Sequence 8, Appl1
45	95	2.2	4303	2	US-08-460-751-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-276-967-2
Sequence 2, Application US/08276967
Patent No. 5851817
GENERAL INFORMATION:
APPLICANT: Hardy, Daniel M.
TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
TITLE OF INVENTION: Sperm
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSER: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,967
FILING DATE: Submitted Herewith
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: UTSD:418\KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2476 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-276-967-2
Query Match 2.7%, Score 117, DB 2, Length 2476;
Best local Similarity 21.6%, Pred. No. 0.19;
Matches 94; Conservative 53; Mismatches 140; Indels 148; Gaps 23;
OY 484 VASAIISODLYFSPFGSGIKQIVKON-----ISVTLRFAP- 522
DB 85 VLGSIRKHLIFSGQ--PGRSMQPVSVNYTSQGIQFTLVGVFKIRPRAVDAISIARC 142
```

```

QY 523 --SFQO-----EAS-----RGLTVSFIPIFKEGVETVPTDTSK-VYL 559
Db 143 EESFPQCFEDNAHPCDWMQASODGGMROGKNKFTL--QAPGPGISLNGHYIFL 199
QY 560 RTPMWDGLPSLTISVSNISVPRQVACLTFEKRSGVVCOTGAFAIMOQRRAEIEF 619
Db 200 ETDKFSQAGOSFRLVSPFCAP-AVICVFETHMYGL-----GQGRKRLILL 245
QY 620 -----SLDEDVLPK-PSFHHSFWVNISNCSPTSGKO-LDLFSVTLPTVDTLV 668
Db 246 GSPAGSPSSLMERVGQSPK-----WLNSTVIPSQHQQPMLIFE----- 287
QY 669 ILAAGVGVLISALDILICCVAKKKKTKNGKPAVGIVNGNINTEMPROPKKEQOKRKD 728
Db 288 ---AVRGNTAFVVALGFLV-----INHGTCRGPSSETSVSTKRPVAPTE---KPT 331
QY 729 NDSHVAV-IEDIMVYGHLLDSSGSLQPEVDIYRPFQGMGVCPSPTICRAPAK 787
Db 332 VPSEIYTIPEKPMVH-----MEKPIVHTEKP---TVPEKPTIPEKSTVPTK 378
QY 788 -----LATEEP-----PRSPSESESE-----YTFSPNNGDVSSKDTDIP- 824
Db 379 PTVKEPTLPPEGPTVAERTTPREGPAVPKGPVLTETPMSHTKSTVHTKPIPLPT 438
QY 825 ---LNTQEPMEPAE 836
Db 439 GKSTIPEKPMVPTK 453

```

RESULT 2

```

US-08-755-587-189
; Sequence 189, Application US/08755587
; Patent No. 6045997

```

GENERAL INFORMATION:

```

; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Asmworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC

```

COUNTRY: USA

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sidley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 189:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1589 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-189
Query Match
Best Local Similarity 17.7%; Score 116; DB 3; Length 1589;
Matches 134; Conservative 105; Mismatches 233; Indels 286; Gaps 34;

```

```

QY 198 LPWFHPRNVSQF-----SIANRSSIKRCLIESVEEGESATLMSANVPEGFPDEL 249
Db 558 LPNINENEPGFCALTKLISVNEALRKAMKLFSDIENSEPEPSAKVG---PREFSS-- 612
QY 250 MTQWQVPAHLRASVSFLNRLNSCERKEE-----VEYIPGST-----INPEYFKL 297
Db 613 -----AHHDSVASYAFKIKKONTKESDKSSKQVTLQNNIEMTTCIFGRNPEKRYK 664
QY 298 EDKQGMAGNFNLGCGDQDQSPILRLQFVLVQHQNESKILYV---VDLSNER 353
Db 665 NTKHEDSYTSQRNNLNSGSMSS-----TSGPYIHKGSDL--- 703
QY 354 AMSLTIEPRVYKSRKFPVPCFVCLERSCTSNLTLTSGSKHKISFLCDDLTRLMMVER 413
Db 704 -----PADQSK-----CPESCQYARENTQ-IKENIS---DLTCLIMAAE 742
QY 414 TISCTDHRICOKRSYSLQVPSDILHPELHDSW----- 448
Db 743 T-----CMKSSDKKQLPDSKMEQNIKEFNISFQASGKNTRVSKESLNKSVNIENRE 794
QY 449 --KLVPKDRL-SLVYPAQKLOOH--HEKPCNTSPSYLVASIPQDLYFGSFCPGS 503
Db 795 TDELTVISDSLSNKLILGINKDKMHTSCHK-----ALSTKKYFEDH--PIYV 840
QY 504 IKQIQVKNISVTLRTPAPSPQDEASQGLTVSE-----IPYKPE--BGVFTVPTDTK 554
Db 841 VSQLPQAQH-----PEYEISTKEPILLSHTASGKKVKIMQSLDKVKVLLPDETQ 891
QY 555 SKYLYR-TPMWDKGLPSLTISVSNISVPRQVACLTFEKRSGVVCOTGAFAIMOQR 613
Db 892 ---YVRKTASFQSGSKPL-----KDSKELTLAYEK-----IEVTAS 925
QY 614 RAEI---FSLDEDVLPKPSFHHSFWVNISNCSPTSGK-QLDLFSVTLPTVDTLV 669
Db 926 KCEMKNFVSKETEMLQONIMHYROTENLKTSGTSSKVOENIENNVENR----- 978
QY 670 ILAAGVGVLISALDILICCVAKKKKTKNGKPAVGIVNGNINTEMPROPKKEQOKRK 727
Db 979 -----ICICQSSYPVEDSALAYTDEDSRKTQVRE--SSLSKGRKL 1019
QY 728 -----DNDSHYAVI-----EDTM 741
Db 1020 REQDKLGTNTIKIECKEHTEDFAGNASYEHSLVITRTEIDTNHVSSENOVSTLLSDPNV 1079
QY 742 VYGHLLDSS-----SGSFLOPEVDIYRPFQGMGVCPSPTICRAPAKLATE 791
Db 1080 CHSTLSQSSFCCHCDMDHNDGIFLKNKIDS-----DVPPDM---KNAEGNTI 1123
QY 792 EPPRSPSESESEPTFSH-----PNNGDVSSKDTDI 823
Db 1124 SPVRSATKERNLHQITINVCQKLETNTSPHANNDAI 1161

```

RESULT 3

```

US-08-459-568-4
; Sequence 4, Application US/08459568
; Patent No. 5811304

```

GENERAL INFORMATION:

```

; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700

```

```

Query Match Summary      2.5%: Score 110.5; DB 2; Length 17719;
Best Local Similarity    19.1%: Pred. No.0.46;
Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42.

QY 78 FTFCSCSEPMNFV---TEIQKIDCMGSCPFGVQLOLPSTSLPTLRFTIMDVKANR 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 FMFPCOHCERKFTTQKQGLERHMH-ISTVNAH-----CKK 393
QY - 134 SIGELQCSIPRLR-----IGPSCSCDGYT-NSISGRDATVVRIGT 176
Db 394 YCGKAFGTQINRRHRERHNEAGLAKRPSOTLOPSEDLDGKASGENAVASKDDSSPPSGP 453
QY 177 FC-----SNGYSRIKMOEG-----VKKALHLPWHPNRVSGF 209
   . ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 DCLINSEKASODTINSSVSENGEVEUKELHPCYCKKVGTHTNNRRHQRYRHEHLIPK 513
QY 210 SIAMNSIKRLCITSVSEEGESATLMSANYPEGPEDE-----LATWQVVAHLRA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 GVRRRGGL-----PPRPAEOAQTVQVYVSTPEEGEADVYIND----- 557
QY 263 SVSFPLNPLSNCKEKEVEEYIIPG-----STPPEVFKLEDKOPGNMAGNFNLQGC 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 ---ISSNIS-----ENLNYIDQKIQTNNTNSCDVIEML----- 589
QY 317 DODAGSPGILRQFOVYLQHPQN-ESNKITYV-DLSNERAMSLTIEPRPVKOSKRFVPGC 374
Db 550 SASADLYGINCLLPVTAETITONIKTTQVQVPEEDLPKEPLGSTNSE--AKRRRASPPA 646
QY 375 FVCLSESR-----CSSNLITLTSGSKHKHISFLCDDLRLTMANNEKTICTSDHRYCQR 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 LPKIAEFDSDPMVPSCSLSLPSTIEVAVSF-----HKEKSV-----YLSKK 690
QY 427 SYSLOVPSDILHLPVELHDFSKLLVPPKDRLSLVLPAPAKLOQHTHEKPCMTSESLVAS 486
Db 691 LKQLQOTQDKLTPPAGISATETAKGP---VCSAPASML-----PVYSSRKRRRTS 739
QY 487 AIPSDQLFYGSCFPGSJKIQLOVKONISVTLRTFA-PASQOEA-SROGLTVSFIPEKKE 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 SEPPSSPGH-----SPALRDFPKPSDGKAAMWDAGLT----- 770
QY 545 GVFYTVTPDTKSAVYLRTPMWDGRGLPSLTLSVSNNIS--VPDQVACLTFPEKSGVVCQTG 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

```

Db 394 YCGAFGTQINRRRHERHHEAGLKRKPSQTLQPSEDLADGKASGENVASKDDSSPSPISGP 453
Qy 177 FC-----SNGTVSRIKMOEG-----VKMALHLPMFHRNVS GF 209
Db 454 DCLIMNSEKASQDITINSVVEENGEVKEHLAPCKYCKKVFGTHTMRRRQRVHERHLPK 513
Qy 210 SIARRSSIKRLCITIESEVEGEGSATLMSANYPGEFPDE-----LMTWQFVYPAHLRA 262
Db 514 GVRKRGGLE-----EPQPPAQOATQNVVYVSTPEEEGEGADVVYIMD----- 557
Qy 263 SVSFLNLSNCKEKEEREVEYIYG-----STNPEVFKLEDQPGMAGNFNLISLQGC 316
Db 558 -----ISSNIS-----ENLNYIDGKIOTNNNTSNCVIEIEME----- 589
Qy 317 DQDASPGILRLQFOVLVQHPON-ESNKIYV-DLSNERAMSLTEPRPVYKSRKFPVGC 374
Db 590 SASADLYGINCLLTPVVEITQNIKTQVPTEDLPKEPLGSTNSE--AKKRITASPA 646
Qy 375 FVCLSESR-----CSSNLITLTSGSKHKISFLCDDLRLMNMVETISCTDHRICOR 426
Db 647 LPKIKAFETSDPMVPSCLSLPLSISTTEAVSF-----HKERSV-----YLSK 690
Qy 427 SYSLOVPSDILHLPEVLHDFSMKLLVPKDRSLVLPQAQKLOQHTHEKPCITSYLVA 486
Db 691 LKOLLQODKLTTPAGISATEIAKLG---VCVSAPASML-----PVTSSRFKRRTS 739
Qy 487 AIPSDLYFGSFCGSGIKQIVQKONISVTLRTFA-PSFOEA-SROGLTVSFTPYKKE 544
Db 740 SPSPSPQH-----SPALRDFGKPSDGAAMTDAGLT----- 770
Qy 545 GVFTVTPDTKSKVYLRTPNMDRGLPSLTSVSNIS--VPRQVACLTFEKKRSQVVCOTG 602
Db 771 -----SKSKLSHSHS-----DSPAMSLSGRDERETVSPCEDEKMKSEKWTAS 813
Qy 603 RAFMIIOEQ-----RTRAEELIFSLD---EDVLPRKPSFHHSFWNINSCPTSQKQ 650
Db 814 SAESVVCQODPLDSSGVKQKAEQGTGVQWESVYL-----DLSYHKHNCDSSECKE 865
Qy 651 LDLLFSVTLPRVTLVYILIAVGGVLLSALGLIICVKKRKKKTN--KSPAVGIV 707
Db 866 FKESHSVQPT-----CSAVKRRKRPYTCMLQGVLLNEY 897
Qy 708 NG-NINTEMPR-----QPKFKQGRKDNDSHYAVIEDTVYGHLLQDSSGSLQPEVD 760
Db 898 NGIDLPEVNPADGTRSPSPCKSLAQDPD-----LQPGSG-PPAPVYE 940
Qy 761 TYRPFQGTMGVCPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPDVCPSPPALQTPSLSSGQLPRLIPTDSSPPCPPLVTVAITPPPLPTV 994
Qy 795 PRSPSESESEPTTFSPNNGDVSSKDTDIPLN-----TOEPMEPAE 836
Db 995 PLPAPSSASBPHPCPSPPLSN--ATAQSPILPISPTVSPSPPIPPVE 1039

```

RESULT 5
 US-08-516-859a-4
 Sequence 4, Application US/08516859A
 Patent No. 6069231

GENERAL INFORMATION:
 APPLICANT: Huang, Shi
 TITLE OF INVENTION: Retinoblastoma Protein - Interacting
 TITLE OF INVENTION: Zinc Finger Proteins
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell & Flores LLP
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEO ID NO: 4
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-516-859a-4

```

Query Match 2.5%; Score 110.5; DB 3; Length 1719;
 Best Local Similarity 19.1%; Pred. No. 0.46; Mismatches 289; Indels 333; Gaps 42;

```

Qy 78 FTFCSCSPENHFV-----IELOKNIDCMGSPCFGEVOLQSTSLPLNRTFINDVNAH 133
Db 358 FMFCQCHERKFTYKQGERMHNIH-----ISVYNAHF-----KCK 393
Qy 134 STGLEQSLPRLQ-----IGPESCPDGYT--HSISGRIDATVVRIGT 176
Db 394 YCGAFGTQINRRRHERHHEAGLKRKPSQTLQPSEDLADGKASGENVASKDDSSPSPISGP 453
Qy 177 FC-----SNGTVSRIKMOEG-----VKMALHLPMFHRNVS GF 209
Db 454 DCLIMNSEKASQDITINSVVEENGEVKEHLAPCKYCKKVFGTHTMRRRQRVHERHLPK 513
Qy 210 SIARRSSIKRLCITIESEVEGEGSATLMSANYPGEFPDE-----LMTWQFVYPAHLRA 262
Db 514 GVRKRGGLE-----EPQPPAQOATQNVVYVSTPEEEGEGADVVYIMD----- 557
Qy 263 SVSFLNLSNCKEKEEREVEYIYG-----STNPEVFKLEDQPGMAGNFNLISLQGC 316
Db 558 -----ISSNIS-----ENLNYIDGKIOTNNNTSNCVIEIEME----- 589
Qy 317 DQDASPGILRLQFOVLVQHPON-ESNKIYV-DLSNERAMSLTEPRPVYKSRKFPVGC 374
Db 590 SASADLYGINCLLTPVVEITQNIKTQVPTEDLPKEPLGSTNSE--AKKRITASPA 646
Qy 375 FVCLSESR-----CSSNLITLTSGSKHKISFLCDDLRLMNMVETISCTDHRICOR 426
Db 647 LPKIKAFETSDPMVPSCLSLPLSISTTEAVSF-----HKERSV-----YLSK 690
Qy 427 SYSLOVPSDILHLPEVLHDFSMKLLVPKDRSLVLPQAQKLOQHTHEKPCITSYLVA 486
Db 691 LKOLLQODKLTTPAGISATEIAKLG---VCVSAPASML-----PVTSSRFKRRTS 739
Qy 487 AIPSDLYFGSFCGSGIKQIVQKONISVTLRTFA-PSFOEA-SROGLTVSFTPYKKE 544
Db 740 SPSPSPQH-----SPALRDFGKPSDGAAMTDAGLT----- 770
Qy 545 GVFTVTPDTKSKVYLRTPNMDRGLPSLTSVSNIS--VPRQVACLTFEKKRSQVVCOTG 602
Db 771 -----SKSKLSHSHS-----DSPAMSLSGRDERETVSPCEDEKMKSEKWTAS 813
Qy 603 RAFMIIOEQ-----RTRAEELIFSLD---EDVLPRKPSFHHSFWNINSCPTSQKQ 650

```

```
Db 814 SASSVCNQOPLDSSGVKKAEGTGTQVQMSVLT-----DLSVHKHKHSDSGKE 865
Qy 651 LDLESVTLTPRTVDLTVIIIAVGGVLLSALGLIICVKKKKKTN--KGPVGIY 707
Db 866 FKESHVQPT-----CSAVKKRKPTTCLQKVLINEX 897
Qy 708 NG-NINTEMPT-----QPKKFKGRKNDSHYAVIEDIMYVGHLLQDSSGSFLOPEVD 760
Db 898 NGIDLPEVNADGTRSPSPCKSLAQPPD-----LGPSSG-FPAFTVE 940
Qy 761 TYRFGQTMGVCPPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPVCPSSPALQTPSLSSGQLPPLLTDPDSSPPPCPPVLTATPPPLPTV 994
Qy 795 PRSPSESEPTFFSHPNNGDVSSKDTDIPLN-----TQPMPEAE 836
Db 995 PLPAPSSASPHPCPSPLSN--ATAQSPILPILSTVSPSPSPPIPVPE 1039
```

```
RESULT 6
US-09-586-472-4
; Sequence 4, Application US/09586472
; Patent No. 633335
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/586,472
; FILING DATE: 01-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/528,706
; FILING DATE: 17-MAR-2000
; APPLICATION NUMBER: US 08/516,859
; FILING DATE: 18-AUG-1995
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 4130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-586-472-4
```

```
Query Match 2.5%; Score 110.5; DB 4; Length 1719;
Best local similarity 19.1%; Pred. No. 0.46;
Matches 169; Conservative 96; Mismatches 289; Indels 333; Gaps 42;
```

```
Qy 78 FTSCQSPENHFV-----IEIQKNIDCMGSPCPCEVQLOPSTLPLNRTFTIMDVKAHK 133
Db 358 FMPCQCHCKERFTTKQGLERHMHIT-----ISTVNAF-----KCK 393
Qy 134 SIGLEQFSIPRLQ-----IGPESCPDGYT--HSISGRIDATVVRIGT 176
Db 394 YCGKAFQTOINRRHERHERHAGLKRKPSQTLQPSBEDLADKASENVAASDDSSPSLGP 453
Qy 177 FC-----SNGTVSRIKMOEG-----YKMLHLFWFHRNVSGF 209
Db 454 DCLIMNEKASOPTINSVVEENGEVKEHLPCCKYCKVFQHTMRHRQVRHERHLIPK 513
Qy 210 SIARRSIRKLCITESTFEGEGSATLMSANTPEEPFDE-----LMTQFVYPALIRA 262
Db 514 GVRKRGLE-----EPQPAQAOATQNYVYVSTPEEEDVADVYIMD----- 557
Qy 263 SVSFLNFNLNCSCKEREEVEYIIPG-----STNPVYFKLEDKQPGNMAGNFNLSTQGC 316
Db 558 -----ISSNIS-----ENLNTYIDGKIQTNNNTSNCVIEHE----- 589
Qy 317 DQDAQSPGILRQFVLVQHPQN-ESNKIYV-DLSNERAMSLTIEPRYKQSKRYVGC 374
Db 590 SASADLYGINCLLTPVTEITQNIKTQVPYTEDLPKEPLGSTNSE--AKKRTASPPA 646
Qy 375 FVCLSEST-----CSSNLITLSSGSKHKISFLCDLDTLRMMNVERTISCTDHRVYQK 426
Db 647 LPRKIAETSDPMVPSLSLPLSISTEAVSF-----HKESV-----YLSK 690
Qy 427 SYSLOVPSDILHLPELCHDFESWKLVPKDRSLVLPVPAOKLQOHTHKPCNTSPSYLVA 486
Db 691 LKQLOLQODKLTTPAGISAEIATLGP-----VCYSASASML-----PTSSKFKRTS 739
Qy 487 AIPSQDLVFGSFCPGGSIKQIQVKNISVTLRTFA-PSFOQA-SROGLTVSFILPYREE 544
Db 740 SPSPSPQH-----SPALRDEGKPSDGAAMTDAGLT----- 770
Qy 545 GVFTVTPDTKSKYVLTLPNMDRGLPSLTVSWMNS--VPRQVACLFFPKERSGVVQOTG 602
Db 771 -----SKRSKLESHS-----DSPAWSLSGHDERETVSPCFDEKMKSEMTAS 813
Qy 603 RAFWIIQEO-----RTFAEIFSLD--EDVLKPSPHHNSFWNINSPTSQKQ 650
Db 814 SASSVCNQOPLDSSGVKKAEGTGTQVQMSVLT-----DLSVHKHKHSDSGKE 865
Qy 651 LDLESVTLTPRTVDLTVIIIAVGGVLLSALGLIICVKKKKKTN--KGPVGIY 707
Db 866 FKESHVQPT-----CSAVKKRKPTTCLQKVLINEX 897
Qy 708 NG-NINTEMPT-----QPKKFKGRKNDSHYAVIEDIMYVGHLLQDSSGSFLOPEVD 760
Db 898 NGIDLPEVNADGTRSPSPCKSLAQPPD-----LGPSSG-FPAFTVE 940
Qy 761 TYRFGQTMGVCPPSP-----PTICSR-----APTAKLATEEPP----- 794
Db 941 S-----TPVCPSSPALQTPSLSSGQLPPLLTDPDSSPPPCPPVLTATPPPLPTV 994
Qy 795 PRSPSESEPTFFSHPNNGDVSSKDTDIPLN-----TQPMPEAE 836
Db 995 PLPAPSSASPHPCPSPLSN--ATAQSPILPILSTVSPSPSPPIPVPE 1039
```

```
RESULT 7
US-09-528-706-4
; Sequence 4, Application US/09528706
; Patent No. 6468985
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
```


PAGE

DY
300 KOPGMMG-----NPNLSLQ-----GCCDQAOSPGILRLQOVUVOHQ-----NESKTI--- 344
|||::|::|::|::|::|::|::|::|::|::|
Db 1241 KO--NVAGLEQPPTPILODRSLNDSPHSGLIMHANSFKIREANLEGLOGNOTOWER 1298

OY 345 -----YVDSLNERAMSLTIEPPRVKSRKFVBGCFCVELSRTCSSNLTLTS 391
Db 1299 PFTBTMSSMASNOHVITTOGRKRSLK---QPRLSGOEIKF-----ERVIANDTSTQ 1346

OY 392 GSKHKISFLCDLLRMLNMNEVKITSCTDHRY--CORXSYSLQVPSDILLPELVLDHDSWK 449
|||::|::|::|::|::|::|::|::|::|:
Db 1347 WSKMNMYLAOGTQLQIENEXEKRAITOSPLSDCSMRN-----HTVIOMNSDA-- 1394

OY 450 LLVRPDRLS-----LYLPAPAKILOQTHTEKPCTSTSYLVAISPQSODLFPSFCBG 502
|||::|::|::|::|::|::|::|::|::|:
Db 1395 LPVAKESASPVVRNTDLTKTRISOHNSSHLPASACNYTFRETSQV-----OEG 1442

OY 503 SIKOIQVNKNISVTLTRTPAFSGOEASHROGLTVSEIPFYKEGEVFVVY----PDTKSVY 558
1443 S-----HFLOEAKRNINLSIAFVLGITIGEGOKFSLSLKSAINTQPMY 1483

OY 559 LRPFWMDRGLESLSVSNNISVPRDOVACLFFPKERSGVQCQTGRAFMIOBORTAEFI 618
Db 1484 KKLETUYLLQGLESETS-----DKVELLS-----QV 1509

OY 619 FSLDEVDLRKSFHHSHSWWNINSOSPYSKGOLDLFEVTLTPRTVDTLYLIAAVGGCV 678
Db 1510 HVDED-----SEPPTSNDSP---GHIDL-----GKI 1535

OY 679 LLLSALGIICCYKKKKKTN-----KGPA--VGJ-YNGNINTEMPRO 718
Db 1536 FLQTCQGIV-----KMKTNPCKGVPRFLMKATSESSEKIFPKLCGLVAMDNHYDTQITSE 1589

OY 719 PKRFQKGRKDNDHAYAVIEDTMV-YGHLLODSSGSFLQPEVDYURPOGTMGVCPRSPP 777
Db 1590 EMKSQKKSQWTAFER--KDTILPLGPCENNDDSTAINEGD--KPCREAMMAKOGERG 1644

OY 778 TICCRAPTAKLATERPFRSPRESSESEPTYTSHPNGVSSKDDI 823
Db 1645 RLCSONPPVSKNHOREITVTTLQPEDEKFEYDDPFISIEMKREDFDI 1690

RESULT 9
US-08-473-750-11
; Sequence 11, Application US/08473750
; Patent No. 5834187
; Patent No. 5834187 5786143
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
TITLE OF INVENTION: Sequence and Analysis of Lkp pIIIn
Patent No. 5834187
Patent No. 5834187 5786143
TITLE OF INVENTION: Structural Gene and the lkp pIII Operon of NO. 5834187 5786143
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,750
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/277,321

```

FILING DATE: 19-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 670 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-473-750-11

Query March                2.4%: Score 106; DB 2; Length 670;
Best Local Similarity      20.7%: Pred. No. 0.25; Indels 130; Gaps 18;
Matches 76; Conservative 49; Mismatches 113;

QY   53 AKPCYIVISKR-HITMLSIKSGERIVTFSCSPENHFV-----IEIQ---KNIDCMGSP 103
    | ||| | : : : : : | : | : | : | : | : | : | : | : | : | : | : |
DB   413 ALPAIVASARNLDFTVHIOSEINAPITNSL-APENNIMIFKSAVELGYSFAGITCTSNP 471
QY   104 CPFGVOL--QPSTSLPTLNRTFIWDYKAHKSTGLQLQSIPRLRIQIGPGESCDDGYTH 161
    || : : : : | : | : | : | : | : | : | : | : | : | : | : |
DB   472 CPTMKLPPLFHFLPTNLTP-----PKKKNSDG--- 497
QY   162 SISGHIDATVRIRGFCNGTYSRIRIKMQGVKMAHLPMFHPNRNVSGSIANRRSIRKLC 221
    : : : : : : : : : | ||| | : : : | : | : | : | : | : | : | : |
DB   498 -----GEIFLHNESING-----VSFOIGVTNTSLMVANKN-----NFSSLKVLAM 539
QY   222 IIESVEEGSATLMSANYPEGPEDELMTWOVVPAHLRASVSFL-NFNLSNCRKKEER 280
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   540 V---PFNSKSISTL-----HRAKFHLTDPFSSLAND----- 568
QY   281 VEYIYGSTTDEYFKLEDKOPGNAGNFNISLGCCDDADOSPGLRLQFOYLVOYPONE 340
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   569 ----TTIDPMNTSIKIMLETWRGSTGNFSVKYGEDK----GDISIFE----- 609
QY   341 SNKIIVDLSNERMSLT---IEPRPVKOSKRFFVGCEVCLESRCSSMLILTSGRKHKI 397
    : : : : : : : : : | : | : | : | : | : | : | : | : | : | : |
DB   610 -NTPKIIILKKOQRCTLNNAPVSPNVK-----LRAYKKRELEAQSEMEGGTFLRV 659
QY   398 SFLCDDLTL 405
    : | : |
DB   660 N--CDNTT 665

RESULT 10
US-08-477-326-11
Sequence 11, Application US/08477326
Patent No. 5968769
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Britton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5968769
TITLE OF INVENTION: Structural Gene and the LKP Pili Operon of No. 5968769typhab
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Mallitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

RESULT 11
US-09-213-293D-1
Sequence 1, Application US/09213293D
Patent No. 6384299
GENERAL INFORMATION:
APPLICANT: GUTIERREZ-ARMENTA, CRISANTO
APPLICANT: SANZ-BURGOS, ANDRES P.
APPLICANT: XIE, QI
APPLICANT: LOPEZ, PLULA S.
TITLE OF INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS
FILE REFERENCE: 604-469
CURRENT APPLICATION NUMBER: US/09/213, 293D
CURRENT FILING DATE: 1998-12-14
PRIOR APPLICATION NUMBER: PCT/ES96/00130
PRIOR FILING DATE: 1996-06-13
PRIOR APPLICATION NUMBER: PCT/EP97/03070

RESULT 12
US-08-839-008-2
; Sequence 2, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:

APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-2

Query Match 2.4%; Score 104; DB 2; Length 449;
Best Local Similarity 20.6%; Pred. No. 0.19;
Matches 67; Conservative 41; Mismatches 102; Indels 116; Gaps 15;

151 PGESC-----PPGVTHSISGR-----DATV-----RIGTFC----- 178
59 PNKECAWTTTVPBGQVYSLSFRVFDLELHPACRYDLLEVFAAGTSGQRLGRFCGTPRPA 118
179 -----SNGTVSRIMQEGVKMALHLPMFHPRNYSGFSIANRSSIKRLCITESVEEG---- 229
119 PLVAPENQVTLRMTDEGRGGREFLLWYSGRATSG-----TEHFGCGRL 164
230 EGSATLMSANYPG-PPEDELMTWQFVPAHLRASVFLNFNL-----SNCERKREVEYV- 284
165 KAGGTLTTPWPSDYPBGISCSWHIAPDQVIALTFEFKFDLEPPTYC--RYDSVSVEN 222
285 -----IFGSTNPVEFKLEDKQPGN-----MAGFNLSL- 313
223 GAVSDSRRLGKFCGDAVPESSIS-----EGNELLVQFVSDLSVTADGFSASYKT 272
314 --QGCDQDQSPGILRLQFQVLVQHPQNSNKIYVVDLSNERAMSLTIEPRPVKQSRKEV 371
273 LPRGTAKEGQPGPKRGTETKVKVLPKPSQPE-----KTEESPASAPDAPTKQGR- 324
372 PGCFVCLBSRTGSSNLTLSGSKHKT 397
325 ---TGTLQSNFCASSLVATATVSKMV 347

RESULT 13
US-08-475-844-9
Sequence 9, Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lobanekov, Victor V.
APPLICANT: Neiman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Goodwin, Graham H.
APPLICANT: Filippova, Galina N.
APPLICANT: Collins, Steven J.
TITLE OF INVENTION: CTCF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,844
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,680
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,590
REFERENCE/DOCKET NUMBER: 14538A-11-1
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-844-9

Query Match 2.4%; Score 103.5; DB 2; Length 727;
Best Local Similarity 16.4%; Pred. No. 0.51;
Matches 133; Conservative 99; Mismatches 261; Indels 319; Gaps 32;

166 R1ATVVRIGTFCSSNGTVSR-----IKMQEGVKMALHLPMFHPRNYSGF 209
65 QLDPTLQMKTEVMEGVAPAEAAVDDQIITLQVNMEE-----QPINIGEL 113
210 SIAN-----RSSIKRLCITESVEEGSATLMSAN-----YPEGF----- 244
114 QLVQVVPVTVPVATTSVEEL--QGAYENEVSKEGLAESEPMICHTLPLPEGFQVVKVG 170
245 -----PEDELMTWQ--FVPAHLRASVFLNFSNCRKEEREVEYVIG 287
171 ANGEVETLEOGELEPQODPSPWQKDPDQPPAK-----KTKTKKSKLRTYEEG 218
288 STTNPEVFKLEDKOPGMAGNFNLSLQCDQDQSPGILRLQFQVLVQHPQNSNKIYV 347
219 KVDVDSYDFEEL-----QGEILS----- 238
348 DLSNERAMSLTIEPRPVKQSRKFPVPGCFVC-LBSRTC--SSNL-----TLTSGSKRKISF 399
239 EVNAEKVVGMMKPKPKTKIKKGVKTFQCELCSTYCPRRSNIDRHHKSHTDERRPHK-CH 297
400 LCDDLTR--LMMNVETISCTDHRVQGRKSYSLQVPSDILHLPEVLHDSWMLVLPKDR 456

```

Db 298 LCGRAFRVTLLRNHLNHTGTRPHKCPDCMAFVTSGLVR----- 339
QY 457 LSLVLPVAKLOQTHKEPCNTSFSYLAASIPSDLYFGSCPGSJKOIQVONI--S 514
Db 340 -----HRRYKHTHEP-----FKSCMCYASVEVSKLRHITSH 373
QY 515 VLTLPFAPSGOQASROGLV-----SPIPY-----FKEEG--VFYTPDTKS 555
Db 374 TGEPRFOSLCSYASRDYTKLRMRTHSGEKPECYICHARFTOSGTMKMHILQKHTEN 433
QY 556 KVIJLRPMWMDGLSLNSVSNISVPRDVA-----CLFFEKRSVVOQTGRAFI 608
Db 434 VAKHCPHCDTVIARKSDGLVHLRKQHSYIEQKKCRYCDVAFHER-----YALI 483
QY 609 QEORT-RAEELFSLDE-----DYLPRKPSFHHSFWVINSNCSPTSGKO--LDLFS 656
Db 484 QHOKSHKNEKRFKQDCQDYACROERHMMHKRTHTGEPYACSHODKFRKQOLDHMF 543
QY 657 VLTLPRTVDLTVILAAVGGVLLLSALGLIICVKKKKKTKN-----GPAVG 705
Db 544 RYHDPNFPVPAFV-----CSKCGKFTFRRTMARHADNCAGP--- 580
QY 706 IYNGNINTEMROPKKQKGR-----DNDSHVAVIEDTMYG 744
Db 581 --DVEGNGETKRSKGRKRRKRSKEDSSSENAEPDLDDNEDEEPAVE----- 631
QY 745 HLLDSSGSFLQPEVDYRPFQGTMGVCPSPPTICSHAPAKLATEEPPRSPSESE 804
Db 632 -----IEPE-----PEPOVYTPAPPAKKRRGRPORTNQPKNQ 666
QY 805 PYTSHPNNGDVSSKDDIPLNTQEPMEPAE 836
Db 667 PTAIOVEDQNTGAIENTIVEVKKKEPDAEPAE 698

RESULT 14
PCT-0595-08429-9
; Sequence 9, Application PC/TUS9508429
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CTCF
; NUMBER OF SEQUENCES: 21
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08429
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,680
; FILING DATE: 17-JUN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14538A-11-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 727 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-0595-08429-9
Query Match 2.48; Score 103.5; DB 5; Length 727;
Best Local Similarity 16.48; Pred. No. 0.51;

```

```

Matches 133; Conservative 99; Mismatches 261; Indels 319; Gaps 32;
QY 166 RIDATVVARIGTCFSCNGTVSR-----IMQGVNKAHLPMFHPNVSQF 209
Db 65 QLDPTLLQMKTEVNEGTVAPAEAAVDDQITLITQVMMEE-----QPINIEL 113
QY 210 SIAN-----RSSIKRLCITESVFEFGSATLSAN-----YDEGF 244
Db 114 QLVQVPVAVTVVATTSVEEL---QGAVENTESKEGLAESSEPMICHTLPLDEGFVAVG 170
QY 245 -----PEDELTWQ-----FVYPAHLASVSFLNPNINSCKREPEREYI 287
Db 171 ANGEVETIEGELPEQEDPSNQKDDIOPRAK-----KTKTKSKSLATYEEG 218
QY 288 STNPEVEFKLEDKQPGNAGNFNLQGDODADQSPGILRLFOYLQHPONESNKIYV 347
Db 219 KDQVSVYDFEE-----QOEGLS----- 238
QY 348 DLSNBRASLTIEPRPVKQSKFVPGFCV--LESRTC--SSNL-----TLTSGKHKTSF 399
Db 239 EVNAEKVGNKPKPRTIKKGVKTFQCELCSTYCPRRSNLDRHMSKHTDERPHK-CH 297
QY 400 LCDLDR---LMNVEXTISCTDHRVCOORKSYSLQVPSDILHPVELHDFSMKLLVPR 456
Db 298 LCGRAFRVTLLRNHLNHTGTRPHKCPDCMAFVTSGLVR----- 339
QY 457 LSLVLPVAKLOQTHKEPCNTSFSYLAASIPSDLYFGSCPGSJKOIQVONI--S 514
Db 340 -----HRRYKHTHEP-----FKSCMCYASVEVSKLRHITSH 373
QY 515 VLTLPFAPSGOQASROGLV-----SPIPY-----FKEEG--VFYTPDTKS 555
Db 374 TGEPRFOSLCSYASRDYTKLRMRTHSGEKPECYICHARFTOSGTMKMHILQKHTEN 433
QY 556 KVIJLRPMWMDGLSLNSVSNISVPRDVA-----CLFFEKRSVVOQTGRAFI 608
Db 434 VAKHCPHCDTVIARKSDGLVHLRKQHSYIEQKKCRYCDVAFHER-----YALI 483
QY 609 QEORT-RAEELFSLDE-----DYLPRKPSFHHSFWVINSNCSPTSGKO--LDLFS 656
Db 484 QHOKSHKNEKRFKQDCQDYACROERHMMHKRTHTGEPYACSHODKFRKQOLDHMF 543
QY 657 VLTLPRTVDLTVILAAVGGVLLLSALGLIICVKKKKKTKN-----GPAVG 705
Db 544 RYHDPNFPVPAFV-----CSKCGKFTFRRTMARHADNCAGP--- 580
QY 706 IYNGNINTEMROPKKQKGR-----DNDSHVAVIEDTMYG 744
Db 581 --DVEGNGETKRSKGRKRRKRSKEDSSSENAEPDLDDNEDEEPAVE----- 631
QY 745 HLLDSSGSFLQPEVDYRPFQGTMGVCPSPPTICSHAPAKLATEEPPRSPSESE 804
Db 632 -----IEPE-----PEPOVYTPAPPAKKRRGRPORTNQPKNQ 666
QY 805 PYTSHPNNGDVSSKDDIPLNTQEPMEPAE 836
Db 667 PTAIOVEDQNTGAIENTIVEVKKKEPDAEPAE 698

RESULT 15
US-08-839-008-9
; Sequence 9, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
; APPLICANT: Hurle, Mark R
; APPLICANT: McDonnell, Peter C
; APPLICANT: McNulty, Dean E
; APPLICANT: Rosen, Craig A
; APPLICANT: Siemens, Ivo R
; APPLICANT: Young, Peter R
; APPLICANT: Yue, Tian-Li
; TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
; NUMBER OF SEQUENCES: 9

```


This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:34:02 : Search time 36 Seconds
(without alignments)
334.830 Million cell updates/sec

Title: US-09-899-569A-4

Perfect score: 4394
Sequence: 1 MAGLNCGVSTALLGVLLGA.....SSKDRDIPLNQEMEPAE 836

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications-AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4394	100.0	836	US-09-899-569A-4	Sequence 4, Appli
2	3467	78.9	749	US-09-899-569A-2	Sequence 2, Appli
3	126	2.9	595	US-09-768-826-48	Sequence 48, Appli
4	114	2.6	477	US-09-972-715-9	Sequence 9, Appli
5	110.5	2.5	771	US-09-854-845-47	Sequence 47, Appli
6	110.5	2.5	870	US-09-854-845-31	Sequence 31, Appli
7	110	2.5	766	US-09-854-845-49	Sequence 49, Appli
8	110	2.5	865	US-09-854-845-33	Sequence 33, Appli
9	108.5	2.5	744	US-09-854-845-27	Sequence 43, Appli
10	108.5	2.5	739	US-09-854-845-23	Sequence 27, Appli
11	108	2.5	838	US-09-854-845-29	Sequence 29, Appli
12	108	2.5	683	US-09-770-657-2	Sequence 2, Appli
13	105.5	2.4	683	US-10-023-676-1	Sequence 1, Appli
14	105.5	2.4	1042	US-09-888-615-74	Sequence 74, Appli
15	103.5	2.4	679	US-09-874-198-6	Sequence 6, Appli
16	103.5	2.4	679	US-09-874-238-6	Sequence 6, Appli
17	103	2.3	449	US-09-919-497-89	Sequence 89, Appli
18	103	2.3	458	US-09-925-301-1282	Sequence 1282, Ap
19	103	2.3	458	US-09-925-301-1282	Sequence 1282, Ap

20	99	2.3	994	US-09-852-909-2	Sequence 2, Appli
21	98.5	2.2	1481	US-09-371-900-40	Sequence 40, Appli
22	98.5	2.2	1481	US-09-924-417-60	Sequence 60, Appli
23	97	2.2	556	US-09-888-615-106	Sequence 106, Appli
24	97	2.2	826	US-09-852-909-4	Sequence 4, Appli
25	97	2.2	2026	US-09-901-368-86	Sequence 86, Appli
26	96.5	2.2	937	US-09-949-192-19	Sequence 19, Appli
27	96.5	2.2	1884	US-09-785-770A-17	Sequence 17, Appli
28	96.5	2.2	1907	US-09-785-770A-16	Sequence 16, Appli
29	96	2.2	999	US-09-747-371-2	Sequence 2, Appli
30	95.5	2.2	3594	US-09-911-842-4	Sequence 4, Appli
31	95	2.2	906	US-09-746-491-48	Sequence 48, Appli
32	95	2.2	3034	US-09-737-149-25	Sequence 25, Appli
33	95	2.2	3034	US-09-737-149-30	Sequence 30, Appli
34	94.5	2.2	463	US-09-905-291A-285	Sequence 285, Appli
35	94.5	2.2	463	US-09-909-320-285	Sequence 285, Appli
36	94.5	2.2	463	US-09-909-320-285	Sequence 285, Appli
37	94	2.1	535	US-09-768-826-38	Sequence 38, Appli
38	94	2.1	708	US-09-945-258-16	Sequence 16, Appli
39	94	2.1	811	US-09-992-598-57	Sequence 57, Appli
40	94	2.1	811	US-09-187-368-1	Sequence 1, Appli
41	94	2.1	811	US-09-989-722-57	Sequence 57, Appli
42	94	2.1	811	US-09-989-723-57	Sequence 57, Appli
43	94	2.1	811	US-09-989-727-57	Sequence 57, Appli
44	94	2.1	811	US-09-989-727-57	Sequence 57, Appli
45	94	2.1	811	US-09-989-731-57	Sequence 57, Appli

ALIGNMENTS

RESULT 1: US-09-899-569A-4

Sequence 4, Application US/09899569A

Patent No. US20020142003A1

GENERAL INFORMATION:

APPLICANT: No. US20020142003A1 Albert Schweitzer

APPLICANT: Maria Scherl-Mostageer

APPLICANT: Wolfgang Sommergruber

APPLICANT: Roger Abseher

TITLE OF INVENTION: Tumorspezifisches Antigen (B345)

FILE REFERENCE: 0652.2280001

CURRENT APPLICATION NUMBER: US/09/899,569A

CURRENT FILING DATE: 2001-07-06

PRIOR APPLICATION NUMBER: DE 100 33 080.0

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: DE 101 19 294.0

PRIOR FILING DATE: 2001-04-19

PRIOR APPLICATION NUMBER: US 60/243,158

PRIOR FILING DATE: 2000-10-25

PRIOR APPLICATION NUMBER: US 60/297,747

PRIOR FILING DATE: 2001-06-14

NUMBER OF SEQ ID NOS: 40

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 836

TYPE: PRT

ORGANISM: Homo sapiens

US-09-899-569A-4

Query Match 100.0%; Score 4394; DB 10; Length 836;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLNCGVSTALLGVLLGAARLPKGAFAEIALPRSNITVILIKGTPTILAKPCYIV 60

DB 1 MAGLNCGVSTALLGVLLGAARLPKGAFAEIALPRSNITVILIKGTPTILAKPCYIV 60

QY 61 SKRHITMLSIKSGRIYFTFSCSPENHFIETOKNIDKSGCPFGFVQLQSTSLPT 120

DB 61 SKRHITMLSIKSGRIYFTFSCSPENHFIETOKNIDKSGCPFGFVQLQSTSLPT 120

QY 121 LNRFTMDVAHNSIGLELOFSPRLRQJGPGESCPOGVTHSISGRIDATVATGIFCSN 180

```

|||||
Db 121 LNRFTIDVNAHKSIGLEIFSIPLRLQIGEGSCPDGVHSHISGRIDATVVRIGTFCSN 180
Qy 181 GTVSRIMKMGEGVKALHLPWFHPRNVSGETSIANRSSIKRLCIISVEGEGSATLMSANY 240
Db 181 GTVSRIMKMGEGVKALHLPWFHPRNVSGETSIANRSSIKRLCIISVEGEGSATLMSANY 240
Qy 241 PEGFPEDELTWQFVPAHLRASVSFLNFMNLSNCEKKEEVEYYIPGSTNPEVFKLEDK 300
Db 241 PEGFPEDELTWQFVPAHLRASVSFLNFMNLSNCEKKEEVEYYIPGSTNPEVFKLEDK 300
Qy 301 QPGMAENFMNLSLQGCCDDQASPGILRLQFVQVHQPONESNKIYYVDLSNERAMSLTIE 360
Db 301 QPGMAENFMNLSLQGCCDDQASPGILRLQFVQVHQPONESNKIYYVDLSNERAMSLTIE 360
Qy 361 PRPVQKRKRVPGCFVCELESTCSNLTITSGSKHISFLCDDLTFLMMNVEKTIISCTDH 420
Db 361 PRPVQKRKRVPGCFVCELESTCSNLTITSGSKHISFLCDDLTFLMMNVEKTIISCTDH 420
Qy 421 RYCQKRSYSLQVPSDILHLPVELHDFSMKLLVPKDRSLVLPQAOKLQOHTHEKPCNTSF 480
Db 421 RYCQKRSYSLQVPSDILHLPVELHDFSMKLLVPKDRSLVLPQAOKLQOHTHEKPCNTSF 480
Qy 481 SYLVASAIPOSODLYFGSFCGGSIKQIYQKONISVTLRTAPSFQOASRQGLTVSEIPY 540
Db 481 SYLVASAIPOSODLYFGSFCGGSIKQIYQKONISVTLRTAPSFQOASRQGLTVSEIPY 540
Qy 541 FKREGVTVTPDTKSKYYLRTPMNDRLPSLTSVMNISVPDQVACLTFEKKERSGVYVQ 600
Db 541 FKREGVTVTPDTKSKYYLRTPMNDRLPSLTSVMNISVPDQVACLTFEKKERSGVYVQ 600
Qy 601 TGRAFMIIQORTRAEIEFSLDEDVLPKPSFHHSFVNINSNCSPTSGKQDLLEFVTLT 660
Db 601 TGRAFMIIQORTRAEIEFSLDEDVLPKPSFHHSFVNINSNCSPTSGKQDLLEFVTLT 660
Qy 661 PRYDVLVILIAVGGVLLSALGLITICVKKKKTKNGPAVGINGNINTEMPROPK 720
Db 661 PRYDVLVILIAVGGVLLSALGLITICVKKKKTKNGPAVGINGNINTEMPROPK 720
Qy 721 KFOGRKNDNSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 780
Db 721 KFOGRKNDNSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 780
Qy 781 SRAPTAKLATEEPPRSPSESESEPTFSHPNNGDVSSKDTDIPLNTQEMEPAE 836
Db 781 SRAPTAKLATEEPPRSPSESESEPTFSHPNNGDVSSKDTDIPLNTQEMEPAE 836

```

RESULT 2
US-09-899-569a-2
; Sequence 2, Application US/09899569A
; Patent No. US20020142003A1
; GENERAL INFORMATION:

```

; APPLICANT: No. US20020142003Albert Schweifer
; APPLICANT: Marwa Scherl-Mostageer
; APPLICANT: Wolfgang Sommergruber
; APPLICANT: Roger Abseher
; TITLE OF INVENTION: Tumorsoziozietes Antigen (B345)
; FILE REFERENCE: 0652.228001
; CURRENT APPLICATION NUMBER: US/09/899,569A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: DE 100 33 080.0
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: DE 101 19 294.0
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/243,158
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/297,747
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 749

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-899-569a-2
Query Match 78.9%; Score 3467; DB 10; Length 749;
Best Local Similarity 90.3%; Pred. No. 1.8e-292;
Matches 676; Conservative 9; Mismatches 28; Indels 36; Gaps 6;

```

```

Qy 67 MLSIKSGERIVTFSCSPENHVEVIEIQKINDMSGCPGEGVQLOPSTSLPTLNRTFI 126
Db 1 MLSIKSGERIVTFSCSPENHVEVIEIQKINDMSGCPGEGVQLOPSTSLPTLNRTFI 60
Qy 127 WDVAHKSIGLEIOPFIPRLRQIGPESCPDVTHSISGRIDATVVRIGTFCSGVTSRI 186
Db 61 WDVAHKSIGLEIOPFIPRLRQIGPESCPDVTHSISGRIDATVVRIGTFCSGVTSRI 120
Qy 187 KMGCVKMAHLHPWFHPRNVSGETSIANRSSIKRLCIISVEGEGSATLMSANYPEGPE 246
Db 121 KMGCVKMAHLHPWFHPRNVSGETSIANRSSIKRLCIISVEGEGSATLMSANYPEGPE 180
Qy 247 DELMTWQFVVPAPHLRASVSFLNFMNLSNCEKKEEVEYYIPGSTNPEVFKLEDKQPGNMA 306
Db 181 DELMTWQFVVPAPHLRASVSFLNFMNLSNCEKKEEVEYYIPGSTNPEVFKLEDKQPGNMA 240
Qy 307 GNFNLSLQGCCDDQASPGILRLQFVQVHQPONESNKIYYVDLSNERAMSLTIEPRVQK 366
Db 241 GNFNLSLQGCCDDQASPGILRLQFVQVHQPONESNKIYYVDLSNERAMSLTIEPRVQK 300
Qy 367 SRKVPFCFCVCELESTCSNLTITSGSKHISFLCDDLTFLMMNVEKTIISCTDHRVQK 426
Db 301 SRKVPFCFCVCELESTCSNLTITSGSKHISFLCDDLTFLMMNVEKTIISCTDHRVQK 360
Qy 427 SYSLOVPSDILHLPVELHDFSMKLLVPKDRSLVLPQAOKLQOHTHEKPCNTSFSYLVAS 486
Db 361 SYSLOVPSDILHLPVELHDFSMKLLVPKDRSLVLPQAOKLQOHTHEKPCNTSFSYLVAS 420
Qy 487 AIPSODLYFGSFCGGSIKQIYQKONISVTLRTAPSFQOASRQGLTVSEIPYFKREGV 546
Db 421 AIPSODLYFGSFCGGSIKQIYQKONISVTLRTAPSFQOASRQGLTVSEIPYFKREGV 480
Qy 547 FTVTPDTKSKYYLRTPMNDRLPSLTSVMNISVPDQVACLTFEKKERSGVYVQGTGAFM 606
Db 481 FTVTPDTKSKYYLRTPMNDRLPSLTSVMNISVPDQVACLTFEKKERSGVYVQGTGAFM 540
Qy 607 IIOQRTRAEIEFSLDEDVLPKPSFHHSFVNINSNCSPTSGKQDLLEFVTLPTVYL 666
Db 541 IIOQRTRAEIEFSLDEDVLPKPSFHHSFVNINSNCSPTSGKQDLLEFVTLPTVYL 600
Qy 667 TVTILIAVGGVLLSALGLITICVKKKKTKNGPAVGINGNINTEMPROPKFKOKR 726
Db 601 TVTILIAVGGVLLSALGLITICVKKKKTKNGPAVGINGNINTEMPROPKFKOKR 659
Qy 727 KNDNSHYAVIEDTMVYGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPTIC 786
Db 660 -----EKTMT--PWCQSSRT-----PWT--MGICYRIRAPASCQART 694

```

RESULT 3
US-09-768-826-48
; Sequence 48, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:

```

; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15

```

12 16:03:29

PRIOR APPLICATION NUMBER: 60/148,722
 PENDING DATE: 10/22/2001
 PATENTISM: Homo sapiens
 US-09-768-826-48
 SCORE 126; DB: 10; length 595;
 2.98; Pred. matches 247; Indels 212; gaps 34;
 1.98;

[illegible]

RESULT 4
US-09-972-715-9
Sequence 9, Application US/09972715
; Patent No. US20020110862A1
; GENERAL INFORMATION:
; APPLICANT: Foungias, George
; APPLICANT: Diamandis, Eleftherios
; TITLE OF INVENTION: NOVEL SILEC-LIKE GENE
; FILE REFERENCE 11757 550U1
; CURRENT APPLICATION NUMBER US/09/972,715

1 CURRENT FILING DATE: 2001-10-05
 2
 3 PRIORITY FILING DATE: 2000-10-06
 4
 5 NUMBER OF SEQ ID NOS: 15
 6
 7 SOFTWARE: Patentin version 3.1
 8
 9 SEQ ID NO: 9
 10 LENGTH: 477
 11
 12 TYPE: prt
 13
 14 ORGANISM: Homo sapiens
 15
 16 US-09-972-715-9
 17
 18 2.6% Score 114

[illegible]

```

RESULT 5
US-09-854-845-47 Application US/09854845
; Sequence 47, US20020098491A1
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20020098491A1el Human Semaphorin Homologs and Poly
; FILE REFERENCE: LEX-0177-USA
; CURRENT APPLICATION NUMBER: US/09/854,845
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 47
; LENGTH: 771

```

1/PE PRT
US-09-854-845-47
Best Local Match

Query Match
Best Local Similarity 2.5%, Score 110.5, DB 10, Length 771,
Matches 108, Conservative 83, Mismatches 183, Indels 171, Gaps 25;

DB 403 DLRIMMVEKTI-----SC-TDHRICQKRSYSTLOVPSDILHLVLEHDSKMLVPRD 455
QY 456 RLSLVLPACLOOHTHEKPCNTSPSYLASAIRSODLYGSCPGSGIKOIQKONISV 515
DB 317 GRPLIL-----KRNIRKTH-----LTCITVTPAPPTDILFLGTADGMIHKAUVLGSGMI 368
QY 516 TLTPAPSPQDASRQGLTVSFIPYF-----KEGVFTVTPDKSK-----VYLRT-- 562
DB 369 IEEF--GVRESQSVENLVISLQHSLYGAPSGVIOPLPSSCSRYRSCYDCLARDPC 525
QY 563 NMDRGLPSTSVSWNISVPRDOVACITFEKESGVVCTGAFMIIOEOTRAEFISLD 622
DB 485 DVLLPCDPSMLARLMLLNGSMGLSDGGGRVGVGDLVTPDAOPENHSGNYGCVAEENG 544
QY 649 -KOLDLFSVTLTPRTVD-----VNISNCSPT-SG-----VLRGD 484
DB 545 LRTLLASTSLVTRATPAPAKAPATPGAOIAPDVRLVYLAIALAGLICLILASSLLY 604
QY 689 CCVKKKKK-----LVYLAAGGVLLLSALGLI 688
DB 605 ACLEGRGRKRRKTSLSGRASAGSAVLOLYVSGRALQVHNGSMSPSPAPCVLDGPETR 664
QY 714 ENRPQPKFKGKRDNDSDVYAVIEDTMYTG-----HLODSGSGFLOPEVDYTPFG 763
DB 665 QVLCQPPK-----PCVSHAHMECLSLAGLCCHPHLLVHS-----T 713
QY 768 TNGVCPSPPTTCSRAP-----TAKLATEEP-----RSPSESE 804
DB 709 -LGPQSLRHPIMSSSRAPCGDLFVASTLGTGCGEVRLHNSPPLPSCVALVNOPPHS--- 863
QY 803 PTFPS 809
DB 765 PWSFS 769

RESULT 6
Sequence 845-47
Patent No. US20020098491A1
GENERAL INFORMATION: US/09854845
APPLICANT: Waikie, D. Made
TITLE OF INVENTION: SCOVILLE, JOHN
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: US/09/854,845
CURRENT FILING DATE: 2001-05-14
HUMAN SEMAPHORIN HOMOLOGS AND POLYNUCLEOTIDES

US-09-854-845-47
Best Local Match

Query Match
Best Local Similarity 2.5%, Score 110.5, DB 10, Length 870,
Matches 108, Conservative 83, Mismatches 183, Indels 171, Gaps 25;

DB 403 DLRIMMVEKTI-----SC-TDHRICQKRSYSTLOVPSDILHLVLEHDSKMLVPRD 455
QY 456 RLSLVLPACLOOHTHEKPCNTSPSYLASAIRSODLYGSCPGSGIKOIQKONISV 515
DB 317 GRPLIL-----KRNIRKTH-----LTCITVTPAPPTDILFLGTADGMIHKAUVLGSGMI 368
QY 516 TLTPAPSPQDASRQGLTVSFIPYF-----KEGVFTVTPDKSK-----VYLRT-- 562
DB 369 IEEF--GVRESQSVENLVISLQHSLYGAPSGVIOPLPSSCSRYRSCYDCLARDPC 525
QY 563 NMDRGLPSTSVSWNISVPRDOVACITFEKESGVVCTGAFMIIOEOTRAEFISLD 622
DB 485 DVLLPCDPSMLARLMLLNGSMGLSDGGGRVGVGDLVTPDAOPENHSGNYGCVAEENG 544
QY 649 -KOLDLFSVTLTPRTVD-----VNISNCSPT-SG-----VLRGD 484
DB 545 LRTLLASTSLVTRATPAPAKAPATPGAOIAPDVRLVYLAIALAGLICLILASSLLY 604
QY 689 CCVKKKKK-----LVYLAAGGVLLLSALGLI 688
DB 605 ACLEGRGRKRRKTSLSGRASAGSAVLOLYVSGRALQVHNGSMSPSPAPCVLDGPETR 664
QY 714 ENRPQPKFKGKRDNDSDVYAVIEDTMYTG-----HLODSGSGFLOPEVDYTPFG 763
DB 665 QVLCQPPK-----PCVSHAHMECLSLAGLCCHPHLLVHS-----T 713
QY 768 TNGVCPSPPTTCSRAP-----TAKLATEEP-----RSPSESE 804
DB 709 -LGPQSLRHPIMSSSRAPCGDLFVASTLGTGCGEVRLHNSPPLPSCVALVNOPPHS--- 863
QY 803 PTFPS 809
DB 864 PWSFS 868

RESULT 7
Sequence 845-49
Patent No. US20020098491A1
GENERAL INFORMATION: US/09854845
APPLICANT: Waikie, D. Made
TITLE OF INVENTION: SCOVILLE, JOHN
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: US/09/854,845
CURRENT FILING DATE: 2001-05-14
HUMAN SEMAPHORIN HOMOLOGS AND POLYNUCLEOTIDES

US-09-854-845-49
Best Local Match

Query Match
Best Local Similarity 2.5%, Score 110, DB 10, Length 766,
Matches 108, Conservative 83, Mismatches 183, Indels 171, Gaps 25;


```

Db      357 DGSRRNGRYGGVGPERRPPGSCITDLSRSGYNSSDLPSTVLDF-VKLNPLMARVPVPR 415
QY      456 RLSIVLPAKLOOHTHEKPCNTSFSLVSAIPISODLYFGSFCPGCSIKQIOVKONISV 515
Db      416 GRPILL--RRNRYTH-----LTGPTVTPAPPTDLELTGADGWIKRAVVLGSMHI 467
QY      516 TLRFAPASFQOEASRQGLTVSFPIPYF---KEGVFTVTPEDTKSK-----VYLRTP-- 562
Db      468 IEET--GVFRESQGVENVLIVLSLLOHSILYGAPEGVIDLPSSCSRYSQCDICLLADPVC 525
QY      563 NMWDG---LPSLTVSNINISVPRDOVACLTFEERSGVCVOTGRAFMIIODEPTRAEEIF 619
Db      526 GMDGTHACAATLIATKARTALID-----IERGNRCGESSRDGPPLLTRS--VL 575
QY      620 SLDEDVLP--KPSHHHSFV-----VNISNCSP--SG----- 648
Db      576 RGDVLLPCDPFSLALALMLNLNGSMGLSDGGGRYGVGDLVLTDAQPEHSGNYGAYE 635
QY      649 ---KQDLDFSVLTTRYD-----LYVLLTAAGGVLILLSALG 685
Db      636 ENGLTLLASTSLVRAPTAPAPAPAKAPATGCAQLAPRVLLYVLAIALGLCLIASSL 695
QY      686 LIICCVKKKK-----KTKRPANGYNGNIN----- 712
Db      696 LYLACLRRGRGRRRKYSLGRASRAGSANOLOTVSGRALQVHMGSSPSPAMPVLDGP 755
QY      713 --TEMPRPKKFKQGRKNDNSHYAVIEDTMYG-----HLDDSSGSLQPEVDIYRP 764
Db      756 ETROYLCOPR-----PCVSHAHMECLISAGLCQCPHPLLVHS-----CFIP 799
QY      765 FQGTGWVCPSPPTICSRAP-----TAKLATEP-----P-----RSPPE 801
Db      800 ASG--LGVPSQLPHPIWSSSPACGDLFYKSLGTGPQPEVLHSPPLPCVALYNQPPHS 858
QY      802 ESEPYTFS 809
Db      859 ---PWSFS 863

RESULT 9
US-09-854-845-43
US-09-854-845-43, Application US/09854845
Patent No. US20020098491A1
GENERAL INFORMATION:
APPLICANT: Walke, D. Wade
APPLICANT: Wang, Xiaoming
APPLICANT: Scoville, John
TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleo
FILE REFERENCE: Lex-0177-USA
CURRENT APPLICATION NUMBER: US-09/854,845
CURRENT FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: US 60/205,274
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: US 60/208,893
PRIOR FILING DATE: 2000-06-02
NUMBER OF SEQ ID NOS: 50
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 744
TYPE: PRT
ORGANISM: homo sapiens
US-09-854-845-43

Query Match      2.5%; Score 108.5; DB 10; Length 744;
Best Local Similarity 18.6%; Pred. No. 0.54;
Matches 93; Conservative 71; Mismatches 160; Indels 177; Gaps 20;

QY      403 DLTRLMMNVEKRT-----SC--TDHRYCQAKSYSLQVPSDILHLPEVLHDFSQKLLVPRD 455
Db      258 DGSRRNGRYGGVGPERRPPGSCITDLSRSGYNSSDLPSTVLDF-VKLNPLMARVPVPR 316

```

```

QY 456 RLSLVLPAAKLOOHHEKPCNTSFYLVASAIPSDLYFGSFCPGSIKOIOVKONISV 515
DB 317 GRPLLL---KRNIRYTH---LTGPPVTPPACPTDYLFLGTADGWMHKAVVLGSGMH 368
QY 516 TLRTFAPSFOEASROGLTVSFIPYF---KEGVFTVTPDTRSK-----VYLRTP-- 562
DB 369 IEET--QVRESOSVENLVISLLOHSLYVGAPSGVQLPLSSCSRYSCTDCLANDPYC 426
QY 563 NMDRGLPSLTSVMNISVPRDOVACLTFFKERSGVVCGTGFAMITQORTAEELFSLD 622
DB 427 GMDPGTHACAATTTIANRSGSRTALIODIERGNRCESSRDTGPPPLKTRN--VLRGD 484
QY 623 EDVLP--KSFHHHSFW-----VNINCSPT-SG----- 648
DB 485 DVLLPCDDPSNLARLMLNLSGMSLSDGGYRVGVDDLLVTDQAPESHGNYGCAEENG 544
QY 649 -KOLDLPSVTLTPRTVD-----LTVILIAAVGGVLLLSALGLTI 688
DB 545 LRTLLASYSLVTPATPAPAPAPATPGAO LAPVRLLYVLAIALGGLCLILASSLLYV 604
QY 689 CCVKKKKKTKNGKPAVGIYNGNINTEMPROPKFKGKNDNSHYAVIEDTMYGHLIQ 748
DB 605 ACLREGRR-----GRRKYSLSGRASR----- 625
QY 749 DSSGFLDPEVDYTRPFQGTGVCV-----PSPPTICS 781
DB 626 -AGSAVQ-----LQTVSGQCPGEDEGDDEGAGLEGSCLOIIPGEGAPAP----- 672
QY 782 RAPPAKLATEEPPRPSPESE 802
DB 673 -----PPPPPPPAE 682

RESULT 10
US-09-854-845-27
; Sequence 27, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20020098491A1 Human Semaphorin Homologs and Polynucleotid
; FILE REFERENCE: LEX-0177-USA
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 843
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-27

Query Match 2.5%; Score 108.5; DB 10; Length 843;
Best Local Similarity 18.6%; Pred. No. 0.66;
Matches 93; Conservative 71; Mismatches 160; Indels 177; Gaps 20;

QY 403 DLTRLMMNVEKTI-----SC-TDHRVCOBKSYSLQVPSDILHLPVELHDFSKLLVPRKD 455
DB 357 DGSRRMGRYEGGVPERPSCITDSLSQGYNSODLPSLVLDL-VKLHPLMARPVVPT 415
QY 456 RLSLVLPAAKLOOHHEKPCNTSFYLVASAIPSDLYFGSFCPGSIKOIOVKONISV 515
DB 416 GRPLLL---KRNIRYTH---LTGPPVTPPACPTDYLFLGTADGWMHKAVVLGSGMH 467
QY 516 TLRTFAPSFOEASROGLTVSFIPYF---KEGVFTVTPDTRSK-----VYLRTP-- 562
DB 468 IEET--QVRESOSVENLVISLLOHSLYVGAPSGVQLPLSSCSRYSCTDCLANDPYC 525

```

```

QY 563 NMDRGLPSLTSVMNISVPRDOVACLTFFKERSGVVCGTGFAMITQORTAEELFSLD 622
DB 526 GMDPGTHACAATTTIANRSGSRTALIODIERGNRCESSRDTGPPPLKTRN--VLRGD 583
QY 623 EDVLP--KSFHHHSFW-----VNINCSPT-SG----- 648
DB 584 DVLLPCDDPSNLARLMLNLSGMSLSDGGYRVGVDDLLVTDQAPESHGNYGCAEENG 643
QY 649 -KOLDLPSVTLTPRTVD-----LTVILIAAVGGVLLLSALGLTI 688
DB 644 LRTLLASYSLVTPATPAPAPAPATPGAO LAPVRLLYVLAIALGGLCLILASSLLYV 703
QY 689 CCVKKKKKTKNGKPAVGIYNGNINTEMPROPKFKGKNDNSHYAVIEDTMYGHLIQ 748
DB 704 ACLREGRR-----GRRKYSLSGRASR----- 724
QY 749 DSSGFLDPEVDYTRPFQGTGVCV-----PSPPTICS 781
DB 725 -AGSAVQ-----LQTVSGQCPGEDEGDDEGAGLEGSCLOIIPGEGAPAP----- 771
QY 782 RAPPAKLATEEPPRPSPESE 802
DB 772 -----PPPPPPPAE 781

RESULT 11
US-09-854-845-45
; Sequence 45, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Wang, Xiaoming
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20020098491A1 Human Semaphorin Homologs and Polynucleo
; FILE REFERENCE: LEX-0177-USA
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/205,274
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/208,893
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 739
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-854-845-45

Query Match 2.5%; Score 108; DB 10; Length 739;
Best Local Similarity 18.7%; Pred. No. 0.59;
Matches 94; Conservative 73; Mismatches 149; Indels 188; Gaps 22;

QY 403 DLTRLMMNVEKTI-----SC-TDHRVCOBKSYSLQVPSDILHLPVELHDFSKLLVPRKD 455
DB 258 DGSRRMGRYEGGVPERPSCITDSLSQGYNSODLPSLVLDL-VKLHPLMARPVVPT 316
QY 456 RLSLVLPAAKLOOHHEKPCNTSFYLVASAIPSDLYFGSFCPGSIKOIOVKONISV 515
DB 317 GRPLLL---KRNIRYTH---LTGPPVTPPACPTDYLFLGTADGWMHKAVVLGSGMH 368
QY 516 TLRTFAPSFOEASROGLTVSFIPYF---KEGVFTVTPDTRSK-----VYLRTP-- 562
DB 369 IEET--QVRESOSVENLVISLLOHSLYVGAPSGVQLPLSSCSRYSCTDCLANDPYC 426
QY 563 NMDRGLPSLTSVMNISVPRDOVACLTFFKERSGVVCGTGFAMITQORTAEELF 619
DB 427 GMDPGTHACAATTTIANRSLD-----IERGNRCESSRDTGPPPLKTRN--VL 476
QY 620 SLDVLP--KSFHHHSFW-----VNINCSPT-SG----- 648

```

12-18:03:29 2002

12-18:03:29 2002

QY 649 ---KQDOLLEFVTLPTVD-----LTVLTAAGGVLLSL 685

QY 686 LIIICVKKKKKKTKNGPAGVINGNINTEMPORPKKGRKNDSDHVAIVEDTNYGH 745

DB 597 LYVACLREGRR-----GRRKRYSLGRASR-----PSPT 778

QY 746 LIODSGSFLOPEVDTPRFOCTMGVCP-----PSPT 778

DB 621 ---AGGSANVQ-----LQTVSGOCCEDEDEGAGLESCLOIIPGEGAPAP- 667

QY 779 ICSRAPTAKLATEEPPRSPPESE 802

DB 666 -----PPPPPPPAE 677

RESULT 12

US-09-854-845-29

Sequence 29, Application US/09854845

Patent No. US20020098491A1

GENERAL INFORMATION:

APPLICANT: Wang, D. Wade

APPLICANT: Alexander, J. C.

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: NO. US20020098491A1 Human semaphorin homologs and Polynucleotid

FILE REFERENCE: LEX-0177-USA

CURRENT APPLICATION NUMBER: US/09/854,845

PRIOR FILING DATE: 2001-05-14

PRIOR APPLICATION NUMBER: US 60/205,274

PRIOR FILING DATE: 2000-05-18

PRIOR APPLICATION NUMBER: US 60/208,893

PRIOR FILING DATE: 2000-06-02

NUMBER OF SEQ ID NOS: 50

SOFTWARE: FastSeq for Windows version 4.0

SEQ ID NO 238

LENGTH: 238

TYPE: PRT

ORGANISM: homo sapiens

US-09-854-845-29

Query Match 2.58; Score 108; DB 10; Length 638;

Best Local Similarity 18.78; Pred. No. 0.73; Mismatches 149; Indels 188; Gaps 22;

Matches 94; Conservative 73;

QY 403 DTRLMNMVEKTI-----SC-TDHYCQKSKSYSLQVSPDILHLPVELHDFSKLLVPRD 415

DB 337 DGRKMGVEGVEPRPGSCITDLSRSGYSSODLPSLVDF-VKLHPLMARPVPR 415

QY 456 RSLVLVRAKQLOOHTHEKPCNTSFSYLVASAIPSODLYGSCPCGSGIKOIVONKISY 515

DB 416 GRPLL-----KRWIRTH-----LTCGPTVTPAGFTIDLFLGRADGHIHKAIVLGSMH 467

QY 516 TLRTFAPSEQOEARSGLTGTFIYF-----KEEGFTVTPDYTSK-----VYLRP- 562

DB 468 IET-ovfRESQVENIVLSLHSLYVGPASVIOPLSSGSRYSKCDCLARPYC 525

QY 563 NMDRG---LPSLTSVSMNIVSPROVACLTFFKERSGVVOCOTGAFMIIOEORTAREIF 619

DB 526 GMDPGTHACAAATTANTALIOD-----TERGNRGCSSTSDTGPPLKTRS--VL 575

QY 620 SLDEVDLP--KPSFHHSHW-----VNISNCSPT-SG----- 648

DB 576 RGDVLLPCDOPSLARALWILNGSMGLSDGQGYRNGVGLVYDAQPEHSGYCYAE 635

QY 649 ---KQDOLLEFVTLPTVD-----LTVLTAAGGVLLSL 685

DB 636 ENGLTTLASTSLTVRPATPAKAPATPGAGLAPVRLLYVALAALGGLIASSL 695

QY 686 LIIICVKKKKKKTKNGPAGVINGNINTEMPORPKKGRKNDSDHVAIVEDTNYGH 745

DB 696 LYVACLREGRR-----GRRKRYSLGRASR-----PSPT 778

QY 746 LIODSGSFLOPEVDTPRFOCTMGVCP-----PSPT 778

DB 720 ---AGGSANVQ-----LQTVSGOCCEDEDEGAGLESCLOIIPGEGAPAP- 766

QY 779 ICSRAPTAKLATEEPPRSPPESE 802

DB 767 -----PPPPPPPAE 776

RESULT 13

US-09-770-657-2

Sequence 2, Application US/09770657

Patent No. US20020046416A1

GENERAL INFORMATION:

APPLICANT: Gutierrez-Armenta, Crisanto

APPLICANT: Sanz-Burgos, Andres P.

APPLICANT: Xie, Qi

APPLICANT: Lopez, Paula S.

TITLE OF INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS

FILE REFERENCE: Gutierrez

CURRENT APPLICATION NUMBER: US/09/770,657

PRIOR FILING DATE: 2001-01-29

PRIOR APPLICATION NUMBER: PCT/ES96/00130

PRIOR FILING DATE: 1996-06-13

PRIOR APPLICATION NUMBER: PCT/EP97/03070

PRIOR FILING DATE: 1997-06-12

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 683

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Description of Unknown Organism: protein

US-09-770-657-2

Query Match 2.44; Score 105.5; DB 10; Length 683;

Best Local Similarity 19.38; Pred. No. 0.86;

Matches 146; Conservative 98; Mismatches 250; Indels 261; Gaps 38;

QY 109 VQLOPSTSLPTLN--RTFIWDVKAHKSIGLEQFSIRLRQIGPESGCPDGYHISGR 166

DB 88 VQMPVTSAMTAKMLREVLSLPDKPSKIQ-QFL-----SSCRDLTNVTER 136

QY 167 IDATVVRIGTFCNGTYSRIKMGVKKALH-----LPMFHPRNVSGESIANRSIKRL 220

DB 137 V--STLEAIFPTKSSANR-----GVSGLNCANAFDIPMAEAKVASKLYR-VLEAI 188

QY 221 CITESVEFEGESATIMSNYPEGF-----PDELTMTQFVY---PAHLRASVFLNF 269

DB 189 CRAELONGNVNVLTPLLN--ERFHCILACSADIVLATHKTVIMMPAYLE-STGLTAF 245

QY 270 NLSNCKEEREVEYIIPSTNPEVK-----LEDKOPGNAGNPNLSLQCDODDQASG 324

DB 246 DLSKI-----LENVVRHETLPRKLKHLNLSLEQLLESNAWEGSSLVN----- 290

QY 325 ILRLOFVLVHPO--NENKTIYV-----DLSNRASLTETPPVQSRKRVPG- 373

DB 291 ---SLIVARVSASEINRLGLAEPPPSIDDLVSNQ--NVRIEGLPAPFSKRAAGP 342

QY 374 -----CFVCLSERTC-----SSNLTLSGSKHKSIFLCCDITRLMMN-- 410

DB 343 DDNADRPSPKRSNENRNVVERNLQTPPKQSHVSTV-----LKAACHPLQSTFASPT 397

QY 411 ---VEKTIISCDTHRCQKRSYSLOVSPDILHLPVELHDFSKLLVPRDLSLVPAOK 466

DB 398 VCPNVGNEKCAP-----VTHHFFSKIL-----KLAAIR 427

QY 467 LOOHTHEKPC--WTSPSYLVASAIPSOD--LTFGSF-----CPGSGTKOIQVONKISY 515

Patent No. 60520020133847A1
GENERAL INFORMATION: US/10025606
APPLICANT: GUTIERREZ-ARMENTA, CRISANTO
APPLICANT: SANZ-BURGOS, ANDRES P.
APPLICANT: LOPEZ, PAULA S.
TITLE OF INVENTION: PLANT RETINOBLASTOMA-ASSOCIATED PROTEINS
CURRENT REFERENCE: 604-468
CURRENT APPLICATION NUMBER: US/10/025 676
PRIORITY FILING DATE: 2001-12-26
PRIORITY FILING NUMBER: 09/213,294
PRIORITY FILING DATE: 1996-06-23
PRIORITY FILING NUMBER: PCT/EP97/03070
PRIORITY FILING DATE: 1997-06-12
SOFTWARE: PATENT NO. 60520020133847A1

US-10-025-676-1

Query Match
Best Local

Matches	Similarity	Score	RB protein
146; Conservation	2.48; 19.38	105.5	

109 VQLOPSTSLPTLN--RRETFDQV
 98; Mismatches 250; Length 683;

88 VOMPTANSTAKKEREVSISIPRUIQIBESCPDNCN... Indels 2671 GADS
167 IDATVPICT... Indels 2671 GADS

137 V- SIVLEAFPIKSSA I
SFCSCNGT SVSR IKMOEGVKMAH
SSCDRDTNVAV

136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954

-----PEDELMTWOFWV-----
 -----ERHRC LIACSA-----
 -----VAEASKIYR-----

...SYNOPSIS...
...269...

[illegible][illegible]

324 65020

290

— 100 —

INFORMATION - 373

[illegible]

US-888-615-74
Sequence 74, Application US/09888615
Patent No. US20020064856A1
GENERAL INFORMATION
APPLICATION

APPLICANT: PLOWMAN, GREGORY
APPLICANT: WHYTE, DAVID
APPLICANT:

APPLICANT: CAENEPEEL, SEAN
APPLICANT: CHARYDCZAK, GLEN
APPLICANT: MANUCCI, DAVID

INVENTOR: SUDARSANAM, SUCHA
TITLE OF INVENTION: METHOD FOR
APPLICANT: SUDARSANAM, GERARD
ATTORNEY: SUDARSANAM, GERARD

CURRENT REFERENCE: 038602/1214
CURRENT APPLICATION NUMBER:
CURRENT FILING DATE:

PRIOR FILING DATE: 2001-06-26
PRIOR APPLICATION NUMBER: 60/214,000
PRIORITY DATE: 06/21/2001

NUMBER OF SEQ ID NOS: 150
SOFTWARE: PatentIn Ver 3.0
SEQ ID NO: 74

```

NO 74
LENGTH: 1042
TYPE: PRM

```

ORGANISM: Homo sapiens
9-888-615-74

Very Match

	Local Similarity	Score
152; Conservatism	2.48;	104
	18.48;	pred. No

9 STALLGVLLEGAARI, POCNA, ...

ALAVLSHMLTSTFOHSP-1 FAX

68 LSIKSGRIVTEECOCDDNURIT

-----SCQSPENHFVLEIQNIDC
|::: | |

-----FLVQJFELHC

```
OY 117 LLPTLNTFTIMDKAHKSIQLELQFISPRLOIGPESCPOGVTNHSIGRIDATVVRIGT 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 412 KIKLLNOSAMTOSQSN-----SLASCLSKLS--GKSETGKTG-----LINLGN 452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 177 FC-SNGTVSRIMQEG-----VKMALHL-----198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 453 TCYMSNVIOALFMATDFRQVLSLNLNGCNSLMKKLOHLFAFLAHQREAVAPRIPEEAS 512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 199 --PMFHPRNVSGFSIANRSGIKRL-----CIESEFEGEGSATLM 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 513 RPPWETPRQODCSEYLRFLDLHEEEKILKVOASHKPESEILCESETSIQEVASKAAVL 572
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 237 SANYPEGEFDELMTMOFVPAHLRASVFLNENLNCERKEBERVEY-----284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 573 TETPRTSGEKTLEKMF--GGLKLRTHIRCL-----NCRSTQKVEAFDTLSLACPPSS 625
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 285 -----IPGSTTNPEVFK-----LEDKOPGNMAGNF 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 626 LENMSVQDPASSPSIQDGLMQASVGPSEEPVYNPTAATCDLSLVNEKTIQSPNPF 685
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 310 NLSLQCCDDAOSPGILRLQFOVLVOHPONESNKIYVDLSNERAMSLTIEPPVKQSRK 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 686 Y-----CSEMTSV-----NESNKI-----LVNKDVPQKPGGE 713
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 370 FVPGCFVCLSESRTCSSNLTLSGSKHKISFLCDDLRLMANYEKTISCT-----418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 714 TTPSVTDLNLYFLAPEILT--GDN--OYECNCASL--QNAEKTMOITEPEPYLITLL 766
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 419 ----DHRVCOARKSY--SLQVPSDILHPYE-LHDF-----SWKLLVPKRDLSLVLPQAK 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 767 RRSYDQKHVRRKILDNVSLPL-VLELPPVKRITSPSSLSSEMSVDYDFTDISENL--AKK 823
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 467 LO-QHTHEKPCNTSFSYL-----VASAIPSDLYFEGFCPGGSIKOIOVKONISVTLRTF 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 824 LKPSGTDEASCTKLVPYLLSVYVHSGISSESGHYSYA-----RNTSTDSY 872
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 521 -----APS--FOQEASROGLT-----VSFLPYFKEGV 546
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 873 OMVHOSSEALALASSOSHLGRDPSKAVFEQDLNKKEMSKEMFLFNDSRVTFTSFQSVQKI 932
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 547 FTVTP-DTKSKVYLRTPMMDRGLP--SLTSVSMNISVPRQVACLTFFEKERSGVVQOTGR 603
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 933 TSRFPKDTAVYVLLYKKQHSTNGLSGNNPISGLMINDPPLQ-----KELMDAITKQNK 985
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 604 AFMTIOEQRTAAEELFSLDEDEVLPKPSFHHSFWVN--ISNCSPTSG 648
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 986 LYLQEOELNARARALQASASC-----SFRPNGFDNDPDPGCGPTGG 1028
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: November 11, 2002, 11:36:33
Job time : 38 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 11:24:46 ; Search time 39 seconds
(without alignments)
2060.728 Million cell updates/sec

Title: US-09-899-569A-4

Sequence: 1 MAGLNCGVSTALGVLLGA.....SSKDDIPPLNTQEPMEPAE 836
Perfect score: 4394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137.5	3.1	3623	2 T08618	intrinsic factor-B
2	123.5	2.8	986	2 T33135	hypothetical prote
3	118	2.7	2499	1 A30788	mannose 6-phosphat
4	118	2.7	3623	2 T09456	intrinsic factor-B
5	117.5	2.7	1464	2 S58984	development protei
6	117	2.7	603	2 T24315	hypothetical prote
7	117	2.7	2476	2 T34022	zonadhesin - pig
8	116.5	2.7	1027	2 T19173	hypothetical prote
9	115	2.6	582	1 VCVAR	env polyprotein -
10	114	2.6	1228	2 S46754	hypothetical prote
11	113	2.6	1224	2 T40765	web1 protein homol
12	113	2.6	3329	2 T42205	breast cancer susc
13	113	2.6	3329	2 T30904	breast cancer tumo
14	112.5	2.6	440	2 A39613	oligodendrocyte-my
15	112.5	2.6	818	2 S37078	probable protein k
16	112.5	2.6	1185	2 T19212	hypothetical prote
17	112	2.5	1694	2 S50065	siadhesin - mou
18	111.5	2.5	1263	2 AH2011	heterocyst glycoll
19	110.5	2.5	942	1 J01674	protein kinase TMK
20	110.5	2.5	1721	1 T38902	retinoblastoma bin
21	110	2.5	2180	2 A47651	zinc-finger protei
22	109.5	2.5	692	1 S59833	DNA-directed DNA p
23	109.5	2.5	737	2 T31349	hypothetical prote
24	109	2.5	4919	2 T31105	hypothetical prote
25	108.5	2.5	1828	2 A40115	microtubule-associ
26	108	2.5	699	1 I54763	Ra-reactive factor
27	108	2.5	5762	2 A41819	proline-rich pepit
28	106.5	2.4	1621	2 T15264	hypothetical prote
29	106.5	2.4	2083	2 T42721	CRP-ductin-alpha p

30	106	2.4	737	2 T16737	hypothetical prote
31	106	2.4	830	2 T37973	rad6 nucleotide e
32	106	2.4	1091	1 T10CHNL	neural cell adhesi
33	103.5	2.4	866	2 T01171	GI/S transition co
34	103.5	2.4	913	1 T0CHNR	R-cadherin precurs
35	105.5	2.4	4131	2 T21085	hypothetical prote
36	104.5	2.4	810	1 A33380	interleukin-4 rece
37	104.5	2.4	1272	2 C96637	hypothetical prote
38	104	2.4	1487	2 S15904	alpha-1 proteinase
39	104	2.4	1565	2 AD2135	polyketide synthas
40	103.5	2.4	558	2 T37567	probable NADPH cyl
41	103.5	2.4	727	2 G01792	transcription fact
42	103	2.3	449	2 A55362	procollagen I C-pr
43	103	2.3	2347	1 TVHURS	kinase-related pro
44	102.5	2.3	457	1 RWMSTA	T-cell surface gly
45	102.5	2.3	914	2 B96592	hypothetical prote

ALIGNMENTS

RESULT 1

T08618
Intrinsic factor-B12 receptor CUBILIN precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Aug-2002
C:Accession: T08618
R:Moestrup, S.K.; Kozyrak, R.; Kristiansen, M.; Kaye, J.H.; Rasmussen, H.H.; Braut
J. Biol. Chem. 273, 5235-5242, 1998
A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibody
A:Reference number: Z16459; PMID:98148073; PMID:9478979
A:Accession: T08618
A>Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: mRNA
A:Residues: 1-3623 <MOE>
A:Cross-references: EMBL:AF022247; NID:93834379; PIDN:AAC71661.1; PID:93834380
C:Genetics:
A:Gene: CUBILIN
C:Superfamily: Intrinsic factor-B12 receptor CUBILIN; EGF homology
C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra
F:1-20/Domain: signal sequence #status predicted <Sig>
F:21-3623/Product: Intrinsic factor-B12 receptor CUBILIN #status predicted <Mat>
F:133-164/Domain: EGF homology <EGF1>
F:436-467/Domain: EGF homology <EGF>

Query Match 3.1%; Score 137.5; DB 2; Length 3623;
Best Local Similarity 18.9%; Pred. No. 0.72;
Matches 151; Conservative 77; Mismatches 283; Indels 289; Gaps 34;

QY	70	IKSGRIVFPEFCQSPENHFVIEIKNIDCMGCPGEVLOPSTN--LLPLNRTFI-126	
DB	2306	VSSRERLYLKFRHDDGSSYGFKAISACGTVS-GDSGVIESIGIPLPLPANNVFCQ-2364	
QY	127	MDVKAHKSIGLEQFSIPRLQIGPGECPDGYT---HSISGRIDATVVRIGTFCNGT-182	
DB	2365	WFIRGLPGHYLTLSFEEDFN--QSSPG--CTKDEVEIEMNHTSRV-----LARYCGNST-2415	
QY	183	VSIKIQEGYKALHLPMFPRNVSQFSINRSIKRLCTIEVFGE---GSATLSAN-239	
DB	2416	PSSVDRTSSNV-ASVKFVTDGSSVTAAGFRLQKSS-RQVC-----GDDLHGPTGTFISPN-2467	
QY	240	YPGFPEDELMTOQFVPAHLRASVSPLNFNLN-----C-274	
DB	2468	YFNPMPHARICEMTITVQBEGRIVLTNTLRISTQPCNSEHLVFNIGINSNPLOKLC-2527	
QY	275	ERKEERYEYIPGSTNPEVFKLEQKP-----GNMAGNFNLQGCDDAQS-322	
DB	2528	SRVNVNTEFKSSGNTMK-VVEFTDGSRYPGCTASYSTEDAVCGGLPVSQ--CNFSS-2584	
QY	323	PGI-----LRQFOVLQHPQNSNNTIYVDLSNEAMLTIEPRVKSRRFPVQ-373	
DB	2585	PGYNGIRDVARNIDCEWTLSPNPNRNSISIVLELSTESHQDCT-----2629	

[illegible]

RESULT 2

hypothetical protein C45G7.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C:Accession: J33135
R:Dante, M.; Wamsley, P.
A:Submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid C45G7.
A:Reference number: Z21288
A:Accession: J33135
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-986 <DAN>
A:Cross-references: EMBL:AF067611; PDB:1ACJ183.1; GSPDB:GN00022; CESP:C45G7.5
A:Experimental source: strain Bristol N2; clone C45G7
C:Genetics:
A:Gene: CESP:C45G7.5
A:Map position: 4
A:Introns: 373/3; 466/1; 524/3; 559/3; 722/2; 776/3; 839/2; 969/2
C:Superfamily: Caenorhabditis elegans hypothetical protein C45G7.5

Query Match	2.8%	Score 123.5;	DB 2;	Length 986;
Best Local Similarity	18.3%	Pred. No. 1.1;		
Matches 128;	Conservative 102;	Mismatches 242;	Indels 229;	Gaps 32

```

0Y      127  DDVAHKXKIGLEOFSPR-----LRIGGESECPDG 158
Db      387  FETKNMAVACELMWSVPREKRSQLECRLEKLLQALRENNKRLYLKRMGGRKT-PVK 445
0Y      159  VTHSISGRIDATVVRIGTFCNGSTVSRIRIKMOEGYKMAHLHPHFPRNVSGEISIANRSSIK 218
Db      446  LAES-----STVENISMDRETDQVRDIYQSRYRL-LPYLHKETPEMHS----- 468
0Y      219  RLCTIESVFEEGSATILMSANY-----PEGFPEP-----ELMTQFVVPAAHLRASVSFLNF 269
Db      489  QYWTLODLT---EKSDVDESPMYPIQPNQKFLFETPEAEILLITAFNSGNELVYSSTLTKI 545
0Y      270  NLSNCEKEEAV-----EYIIPSTINPEYFKLEDKQPGNMAGNFULSLQGCDQDAQSP 323

```

```

Db      546  RVENNDONHEPEFLPSALRPIFOVPKNTSKPALIGTLTARODFSPRIFYHLLPNCGTBESSD 605
Qy      324  GI-LRLOFOVLVONHO-NESKKITVVDLSNBRAMSLTIEPRPVYKOSRKEVGCFCVLESR 381
Db      606  NENIDAEFGELIYEPKVDPPSKMEAEV-----CFIATSOR 641
Qy      382  TC-SSNLTLTSGSKH-----KISFLCDDTLTLMWVEKTICTDHRVCQRKSYSLQVPSOI 436
Db      642  NIDTSEVEFEFANDSKNFRKVKKEFGD-----SYDSVANSKGINNGFSNGSISSVGDY 694
Qy      437  L---HLPEVL-----HDESKLLVPRKRLSLVLPQA-----KLOQHTHE 473
Db      695  LDRVEIPNNLGAAGADYEOK-----SLNEVPANYLEGRDMISPEGAVELANTNCE 746
Qy      474  KPCNTSFELYVASA---IPSDOLY---FGSPCEGSIKOIQOVON-----ISYTLRF 520
Db      747  -----LVANGKIDTPQGVTAETIGGDGEGYKQHLHNRDKRLATVYLSMRNE 797
Qy      521  APSPQOASRQGLTVSFLPYKEBGEVTPYDTRK---KVYLRTPNMDRGLPSLTYSV 575
Db      798  GANLEK-----FRQILPAIKEDDQKAGKOLEIHFDPEKADRKNSTYSV- 842
Qy      576  MNISPRDQVACLPEFKERSGVVQGTARAMIIQEOCTRAEELFSLDEVLPRSPRHHS 635
Db      843  -----CFYLRQNALIDENQASILLSPNSGHISKI-----HHI 875
Qy      636  FWV-NISNCSP-----TSGKOLDLFSVTLTPRTVDTYLILAAVGGVLLLSALG 685
Db      876  FKVOVDWDCAPKRPATSESTSSSSNSIDIPN-----TLILGVGLTILALL 925
Qy      686  LIICVVK-----KKKTKNK---GAVGYINGNINTEMPROP 719
Db      926  IYVCVSRYQRYLKOKITDOLRCSSASCSYKSPULIPPPP 966

```

RESULT 3

mannose-6-phosphate receptor protein, cation-independent bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999
C:Accession: A25908; A30788; S09404
R:Label, P.: Dahms, N.M.; Kornfeld, S.
J. Biol. Chem. 263, 2563-2570, 1988
A:Title: Cloning and sequence analysis of the cation-independent mannose 6-phosphate
A:Reference number: A92706; MUID:88115411; PMID:2963004
A:Accession: A25908
A:Molecule type: mRNA
A:Residues: 1-249 <IOB>
A:Cross-references: GB:O03527; NID:g162873; PIDN:AAA30455.1; PID:g162874; GB:M15869
R:Glickman, J.N.; Conlbeare, E.; Pearce, B.M.F.
EMBO J. 8, 1041-1047, 1989
A:Title: Specificity of binding of clathrin adaptors to signals on the mannose-6-phos
A:Reference number: S09404; MUID:89305502; PMID:23545438
A:Contents: annotation; HA-II adaptor binding
C:Comment: This protein binds phosphorylated lysosomal enzymes and insulin-like growth
C:Superfamily: mannose 6-phosphate receptor, cation-independent; fibronectin type II
C:Keywords: Golgi apparatus; membrane protein
F:1912-1951/Domain: fibronectin type II repeat homology <2F1>
F:2360-2363/Region: HA-II adaptor binding

Query Match	2.7%	Score 118	DB 1	length 2499
Best Local Similarly	19.9%	Pred. No. 11		
Matches 195	Conservative 100	Mismatches 343	Indels 340	Gaps 56

```

QY 1 MAGNACVSTALLGVLLLGAARLPKGAFAEFLIPRESNTIVYLIKIGPTLL----ARKC 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1710 MHGLACAGATACVCKVPVDP-----PIDIGRAVAPPLINPLIANEVYLNFPSSPFC 1759
QY 57 YIVSRKHIMLSKSSERIVTPESCO-----SPE-----NHVYIEQKNIDC---M 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1760 ---LADHRENTSL-----ITFHCKRGVSKGTPLRLTSCVDFFEFMEETPLVCPDEYK 1806
QY 101 SGCPPEGEVOLQPSLILPLINRFTIWDYKAHK-STGLELQFSIPRLKRGGE-SCPDG 158

```



```
Db 1810 TDGCSLTDEQLYYSPNLSLSLSTFKVTRGPHYISGV-----CTAAAGLDEGGCKRG 1862
Qy 159 VTHISGRDIAVVRIGTFCNSNGTVSRIMQEGVKMALHPHFRPNVSGFSTANSSIK 218
Db 1863 AVCLLSGKSGAFGRLA-----SMKLDLRHDEAVILSLANDCTDP 1903
Qy 219 -----RLCTIISVEFEGESATLMSANYPEGPEDELMQVPAHLVASVFLFNLS 272
Db 1904 PETEDGEPCVFPFVFNKG-----SYEECVESRARLY-----CATTANYDD 1945
Qy 273 N-----CERKEERYEYIIPGTTNPEVFKLEDKOPGNMAGNMLSLGGCDODASPGILRL 328
Db 1946 HEMGFCCKHSTH-----RTSVIIFK-----CEDAD----- 1971
Qy 329 QPQVLVQHQNQNSNKIYVVDLSNERAMSLTIER-----PKQSKRY----- 371
Db 1972 -----VGRPO-----VFSEVRGCEVTPEMKTQVCPPKMECKEYQKHRTYDLRL 2017
Qy 372 ---PGCF-----VCLF-----SRTGSSNLT-----TSGSKHKISF-----LC 401
Db 2018 SLLTSMWSFVHNGASTYINLCKQIKYKPODCSERASVCKSTSGEYOVLGVHTQKLDV 2077
Qy 402 DDLRLMNVETICTDHRVCO-----RKSYS-LQVPSDILHLPVELHDFSM 448
Db 2078 DD--RVITYVYSGKHGCGDKKTSVAVIELCAKTVGRSPFTREDVDSCTYH-----FSW 2128
Qy 449 KLLVRKRLSLVLPVAKIQ--QHTEKPCN--TSESYLVASAIPSODLYFGSCPGSGIK 505
Db 2129 -----DSRAACAVKPOEVQVNGTITNPANGRSFSL-----CDIYFKRFSAGDVR 2174
Qy 506 -----QIOYK-----QNISVTLRTFAPSPFOEASROGLTVSFIFYKEEG-- 545
Db 2175 TNGDRYIEIQLSITGSSSPACSGASICQ--KANDHFGRKVTGTSQTRYYVDGDL 2231
Qy 546 --VFT-----VTPDTKSKYLLR--TPMMDRGLPSLTGSVNINISVPRQVACLFFPKR 594
Db 2232 DVFYSSSSKCGDKKTSVSTIEFHODPLVKDGIPEFS-----HETACQYLFSMH 2282
Qy 595 SCVVCOTGAFAMIOEORTRAEILFSLDEDVLPKPSFHHHSFVNINISNCSPTSGKQLDL 654
Db 2283 TSAVCPPLGAGF-----DEIAGDDAQEHK-----GLSEKSOAVGAVLSL 2322
Qy 655 FSVTLTPRTVDTLVILIAVGGVLLLSALGLIC-----YKKKKKTKNKGPAVITNG 709
Db 2323 ---LVALTACLTLTLTKYKRRKREYMSR--LTNCCRSANYSKSKVKKKEEA---D 2372
Qy 710 NINTE-----MPROPKFKGKRKNDSDHYAV---IEDTM--VYGHLLQDSSGFLQ-P 757
Db 2373 EMETEMLMELIOPRAPRKEGGE--NGHVAAKSVAAADTLSLAHND--EQDSEDEVLTLP 2429
Qy 758 EVDYTRPFO--GTMGVCPSPPTTCSRADTAK-----LATEPPRPSPESESEPTYES 809
Db 2430 EVKVRPPGAPGAEG--GPELRPLPRKAPPLRADRDVGLVRGEPARGRPAAATPISIF 2488
Qy 810 HPNNDGVSSKQDIDPLLN 827
Db 2489 H-----DSDDELH 2498
```

RESULT 4

T09456
Intrinsic factor-B12 receptor Cubillin precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Aug-2002
C:Accession: T09456
R:Kozaki, R.; Kristiansen, M.; Slatkargiu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N
Blood 91, 3593-3600, 1998
A:Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characteriz
ion.
A:Reference number: 216777; MUID:98241400; PMID:9572993
A:Accession: T09456
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-3623 <K02>
A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AAC82612.1; PID:93929529
C:Genetics:
A:Map position: 10p12
C:Superfamily: intrinsic factor-B12 receptor cubillin; EGF homology
C:Keywords: receptor; vitamin B12 uptake
F:1-24/Domain: signal sequence #status predicted <Sig>
F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <Mat>
F:436-467/Domain: EGF homology <EGF>

Query Match 2.7%; Score 118; DB 2; Length 3623;

Best Local Similarity 19.2%; Pred. No. 20;

Matches 160; Conservative 84; Mismatches 271; Indels 320; Gaps 37;

```
Qy 88 HVEYIEIQKINDC-----MSG---PCPFGVOLQPSLTSLPTIKTF--TWYKAHKSIG 136
Db 2352 HPTLPYRDMLFCMHLQSGHYLTISFEDFNQNSG-----CEKDFVEIMDHNTSGNI- 2406
Qy 137 LELQFIPRLRQIGPESCPDGYTHSISGRIDATVYRIGTFCNSNGTVSRIMQEGVKMAL 196
Db 2407 -----LGRYCGNTIPDSIDTSSNTAVR--FVTDGSVT----- 2437
Qy 197 HLPWFRPNRVSGFSIANRSSIKRLCTIISVEFEGES--ATLMSANYPEGPEDELMQVY 255
Db 2438 -----ASGRLRFESSME-----ECGGDLQSGTIGFTSPNYPNPAPHGICMRIT 2483
Qy 256 VPAHLASVSLFNLS---NCEKREERYEYIIPGTTNPEVKL-----EDKOPGN 304
Db 2484 APEGRITLTMFNRLRATHPSG--NNEHYLVFNGIRNSNPQLEKLCSSVAVSNEIKSSG 2541
Qy 305 MA-----GNFNLSSGCDODASPGILRLQFOVLVQHPONESKITVVDLSNE 352
Db 2542 TMKVIFETGDSRPYGGFTASYS--SEDAVCGGLPMTPEKNFTSPYDGVANRNLNCE 2600
Qy 353 RAMSLTIERPPVYKSKRPVGCFLSESRSSNLTLSGSKKHSIFLDCDLRLMANYE 412
Db 2601 WTLG-----NNQGNSSSIHFEDFYLE----- 2623
Qy 413 KTISCTDHRVCOKYSLOVPSDILHLPVELH--FSMKLLVPRKRLSLVLPVAKIQH 470
Db 2624 -----SHQDCQ-----FDVLEFRVADAGPLMWRGSPKPLPLVPIYSQWTH 2668
Qy 471 --THEKPCNTSF--SYLVASAIPSODLYFGSPGGSISQIQVKQINISVTLRTFAFSQ 526
Db 2669 FVTNERVEHIGFAKY-----SFYDCGGIQ----- 2693
Qy 527 EASRQGLTVSFIFYFKEGVFTVTPDTKSKYLLRTPMMDRGLPSLTGSVNINISVPRQV 586
Db 2694 -----IGDSGYIT-----SENYPRAYDSLTHCSMLEAPRGHTI 2727
Qy 587 CLTF--FKERSGVV-----VTLPRTVDTLVILIAVGGVLLLSALGLICVKKKK- 608
Db 2728 TLTFSDFLPEPHYTTCAMDSTVYRANGSPESPITIGXCGNSNPRTIQSGSNQVLVTFNSDH 2787
Qy 609 -----QOQRTRAEILFSLDEDVLPKPSFHHHSFVNINISNCSPTS---GKQLDL 653
Db 2788 SLGGGFYATWNTQTLGCGIIFHSDNGTIRSP--HMPQNPENSRCSWATITKHSKLEI 2845
Qy 654 LFS-----VTLPRTVDTLVILIAVGGVLLLSALGLICVKKKK- 696
Db 2846 SFDNNFLISGGQCCNSFYKVMAGTEVDKALLALGCGNV---APGVYITPSNFTAV 2901
Qy 697 -KTNKGPAVGIV-----NGNINTEMPROPKFKQGRKNDSDHYAVIED-- 739
Db 2902 FQGOEAPAGCFASFSVRCGSNFTGPSYIIS--PMYPRQY-----DNMNCNTYIEAHP 2954
Qy 740 -----TMYGHTLQDS--SGSLQDPVDTYR-----PFGCTMGVCPSPPTI 779
Db 2955 LSVYLLTFVSHLEARSAYTGSCVNDGVHIIKQYSVSTPFAVTCGDEMPAPLTI 3009
```

RESULT 5

S58984
development protein tolkin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C/Accession: S58984
R/Refinelli, A.L.; Xie, T.; Bossie, C.A.; Blackman, R.K.; Padgett, R.W.
Genetics 141, 271-281, 1995
A/Title: The tolkin gene is a tollold/BMP-1 homologue that is essential for Drosophila C
A/Reference number: S58984; MUID:96042912; PMID:8536976
A/Accession: S58984
A/Molecule type: mRNA
A/Residues: 1-1464 <FIN>
A/Cross-references: EMBL:U04777; NID:g1002985; PIDN:AAC47015.1; PID:g1002986
A/Note: the authors did not translate the codon for residue 722
C/Genetics:
A/Gene: tolkin
A/Cross-references: FlyBase:FBgn0004885
C/Superfamily: astacin homology; EGF homology
C/Keywords: hydrolase; metalloproteinase; zinc
F:529-722/Domain: astacin homology <AST>
F:958-993/Domain: EGF homology <EGF>
F:118-1153/Domain: EGF homology <EGF1>
F:614,618,624,673/Binding site: zinc (His, His, His, Tyr) #status predicted
F:615/Active site: Glu #status predicted

Query Match 2.7%; Score 117.5; DB 2; Length 1464;
Best Local Similarity 19.3%; Pred. No. 5.5;
Matches 98; Conservative 63; Mismatches 161; Indels 185; Gaps 24;

QY 105 PFGVYQOPSTSLPTINRFIMVYKAKHSGLELOISIRLRKIGGSCPGCVTSTIS 164
DB 1004 PNGIT-TSPSPPEWYPLKCKCIWETVAPPKHRISLNTFTDL-----EGTAHQOS 1052
QY 165 --GRIDATVY-----RIGTECSNGTYSRIKMGQVKMALHPVHRNV--SGFSI 211
DB 1053 DCGVSTYVSKIGENLKIKIGFCGSSIRPTATSESN--ALRLERHSDKSIORSGFAA 1109
QY 212 ANNSIKR-----LCIIIEYF-----EGEGS-----ATL 235
DB 1110 VFFPDIDECVANNNGCOHECRNTIGSYICMCHNGYSMHENGHDCGECKEYISAPPGTI 1169
QY 236 MSNYPGPFDELMTQGFV-VRPHLRASVFLNPL-SNCKERKEVEVEYIRPSTNPE 293
DB 1170 FSPYRPSTYPMNDCVNHFTTTPGH-RIKLIFNEDESHEOECTYDVNAVY----- 1219
QY 294 VFLEKQKPGNMAGNENLSQGCDDAOSPGIL-----RLQFOVLVQHONESNKIYV 347
DB 1220 -----DGESESSSVLGRFCGDKIIPPI-----STSTNQMYV 1251
QY 348 DLSNERAMSLTIEPRYKQRKVEYGCVCLESFRTGSSNLTLSGHNKISFL-----C 401
DB 1252 -----LKTOKNKKOKNGFTASHSTACGYSFRATSOVOQFYSHAFGNODY 1295
QY 402 DDLRLMANNVEKTICTGDHYCO-----RKSYS-----LQYPSILH----- 438
DB 1296 DD-----GMDEWITIAADNSYVQLILFTPIESSENCTFDYVOVPSIDIDVGOYGPWYG 1351
QY 439 -----LPVELHDFSWKL-----VPRDLRLVLPVPAQKLOQHTH-EKPCNTSFL 483
DB 1352 QYCGNVLPODINSWTHSLVRFKTDGSDVPMKGFASVVAAPNSEGYHSDEDVENSISSE 1411
QY 484 VASAIPIQDLYFGSFCGSGIKQIQVK 510
DB 1412 MVTPFP-----GSLKSIYIE 1426

RESULT 6
T24315
hypothetical protein T01G9.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T24315
R/Lennard, N.

submitted to the EMBL Data Library, July 1996
A/Reference number: Z19874
A/Accession: T24315
A/Status: preliminary; translated from GB/EMBL/DDbJ
A/Molecule type: DNA
A/Residues: 1603 <MIL>
A/Cross-references: EMBL:Z75713; PIDN:CAB00050.1; GSPDB:GN00019; CESP:T01G9.3
A/Experimental source: clone T01G9
C/Genetics:
A/Gene: CESP:T01G9.3
A/Map position: 1
A/Introns: 68/3; 346/1; 407/3; 450/3; 486/3

Query Match 2.7%; Score 117; DB 2; Length 603;
Best Local Similarity 19.9%; Pred. No. 1.6;
Matches 133; Conservative 79; Mismatches 216; Indels 242; Gaps 32;

QY 266 FLNF-----NLSNCKERKEVEYIRPGSTTNEVF-----KLEDK 300
DB 9 FLSEFLVLCISIALPSSCPNLCECDQNDSSWSYCKRAIINDIYAEILNOLPLTLRSLHIQ 68
QY 301 QPGMAG-----NFN-----LSIQCDDQDAOSPGILRLQFOVLVQHPON-----E 340
DB 69 PPSNRIGSNKLRNNDINRPAQLRVLRINCQIPAMRSIRLPSLEVLDLSNNIEHATM 128
QY 341 SN-----KIYVVDLSNERAMSLTIEPRYKQRKVEYGCVCLES-RTCS-SNLTTSGS 393
DB 129 SNFGCMKRLRLDLSNH-----LNLTP-----TGVFTYLALRLSLSNNTSIDLS 175
QY 394 KHKISFL-----CDDLRLMANNVEKTICTDHY---QQRKS-YSLQVPSD 435
DB 176 TNLRLGINSRLVRLDRNPRIPIEHINELFTDY---SOLDELYLNHCNLSITSYLAIDR- 230
QY 436 ILHLVPLHDFSWKLIVPKDRL-----SLVLPVPAKLOQHTHERPCNTSFSYLVAAP 489
DB 231 -----IPQLRLQIGIGNNLKMPTEKLS-----LPQLSVLDLSHNS 267
QY 490 SQDLYFGSPGGSIKQIQVKNT-----SVTLRTFAPSF----- 524
DB 268 IOETTACAF-C-NYISKLDLSHNLGISKSPFEDAFRTMPRLRLDLSFNHMDPDSKW 326
QY 525 -----QQRASQGLVSTPIPFKEGVTYPRDTSKYLLTP-NMDGLP----- 569
DB 327 LGMAQELTSLIALSGNLKNEESWYTL---KSLHLELAYNHIFIPQLPSRYHL 382
QY 570 -----STSVSWNISVPRDQVACLPFKERGVVCGTGRAFMIOEQ----- 611
DB 383 ISLINSNELTYLPDNTILLPNKTPTDITANRHTSHTDIAFLNVEQYVYDGNPWC 442
QY 612 -----RTRAEITFSLDEVDLPKPSF---HHHSFWNISCSPTSGKOLDLP 655
DB 443 SCAIQGLQVHMRDXYARHILINDVNCATPSLVEGSHVLAITDVNCAVLFGARYGL- 500
QY 656 SVTLTPRTVLDLYLIAVGGVLLSALDLIC-CYKAKKKKTKNKPRAVGIYNGNINTE 714
DB 501 -----TQTSMLIIL-----AGVLLFAALLMLIGCTYPLREKQYK-----SYVTR 543
QY 715 MPROPKFKQGRKNDSDHYAVIEDTVYGHLLQDSSGSF---LQPEVDY---RPFQ 767
DB 544 HSKRPL-----TMANHSCSSSINDNHGRLSPFPDFLWSTEFKA 584
QY 768 TMGVCPPSP 777
DB 585 TPPLIPAP 594

RESULT 7
T34022
zonadhesin - pig
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T34022
R/Hardy, D.M.; Garbers, D.L.

J. Biol. Chem. 270, 26025-26028, 1995
 A:Title: A sperm membrane protein that binds in a species-specific manner to the egg ext
 A:Reference number: 221464; MUID:96064658; PMID:7592795
 A:Accession: T34022
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2476 <HAR>
 A:Cross-references: EMBL:U40024; NID:g1066465; PID:g1066466; PIDN:AAC48486.1
 A:Experimental source: strain Melshan; testis
 C:Genetics:
 A:Gene: Zan
 C:Function:
 A:Description: may be involved in sperm adhesion to the zona pellucida

Query Match 2.7%; Score 117; DB 2; Length 2476;
 Best Local Similarity 21.6%; Pred. No. 13;
 Matches 94; Conservative 53; Mismatches 140; Indels 148; Gaps 23;

QY 484 VASAIPODLYFGSCPGSGSIKOIOYKON-----ISVTLRFAP- 522
 DB 85 VLGSTRKHTLFSGQ--PGPSMQPVSVNTSOGIOFTLVGVFGKLEPPAVAVDAISIAFC 142
 QY 523 --SFOQ-----EAS-----ROGLTVSFIPFKEGVFTYPTPKSK-VYL 559
 DB 143 EESFPQCDPEDNAHPCDHWQASODGVMROGKNTFI---QAPGFGISLNGCHYFL 199
 QY 560 RRPMDRGRLPSLTVSMNLSVRDQYACLTFFKRESGVVCOTGRAFMITOEORTAEELF 619
 DB 200 EYDKFSQACOSRFLVSRRPCAP--AVICVTFYHMYGL-----GGCTKLRL 245
 QY 620 -----SLDEDVLPK--PSFHHSHFVNINSCSPYSGKO--LDLFSVTLPRVDTLV 668
 DB 246 GSPAGSPSSLMERNQPOPE-----WNTSVTISGHOQPPQLIFE----- 287
 QY 669 IIAAVGGVLLSALGLIICCVKKKKTKNGKPAVGIYNGINTEMPPROKPKFOGRKD 728
 DB 288 ---AVRGTTAFVVALGFLV-----INHGTCRGPSSETSSTKPVAPFE---KPT 331
 QY 729 NDSHYAV--IEDTMYGHLDDSSGSLQPEVDYRPFQGTACVCPSPPTICSRAPTK 787
 DB 332 VSEIYTTITERPMVH-----MEKPIVTERP---TYPTKPTITERKSTVPTPK 378
 QY 788 -----LATEEP-----PRSPPESEEP-----YTFSHNNQDVSCKDIDP- 824
 DB 379 PTVKEPTLPPEGPTVPARPTTPPEGPAVPKGPVLTLEWPTSHTEKSTVTEKILPT 438
 QY 825 ---LNTQEPMEPAE 836
 DB 439 GKSTIPEKPMPTK 453

RESULT 8

T19173
 hypochelical protein C05D12.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T19173

R:Colles, L.
 Submitted to the EMBL Data Library, August 1995
 A:Reference number: 219084
 A:Accession: T19173

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1027 <WIL>
 A:Cross-references: EMBL:Z50872; PIDN:CAA90754.1; GSPDB:GN00020; CESP:C05D12.2
 A:Experimental source: clone C05D12

C:Genetics:
 A:Gene: CESP:C05D12.2
 A:Map position: 2
 A:Intons: 20/1; 55/2; 128/3; 241/3; 333/1; 352/3; 406/1; 470/2; 493/1; 538/3; 588/1; 66

Query Match 2.7%; Score 116.5; DB 2; Length 1027;
 Best Local Similarity 19.4%; Pred. No. 3.8;

Matches 158; Conservative 90; Mismatches 207; Indels 361; Gaps 44;

QY 142 SIPRLQIGBEGCPGCVTHSISGRIDATYVRIGTCSNGTVSRIMQEVKALHLPWF 201
 DB 338 SIDEATVAEGDSYR---LYSRASKIGDTYT---IFCONGTYYNRM-----PL 380
 QY 202 HPANVSGFSIANSSIKRLCIIESVEFEGESAT-----LMSANYPEGP----- 245
 DB 381 QSKNTLLFNNDMHTD-----VGNMAVIGIEMSTMOCTYFPOMQNSYVRSF 431
 QY 246 EDELMTMOFVPAHLRASVFLNPLSNCKEKEVEYYIPGSTNPEVFKLEKOPGM 305
 DB 432 DCKLREKFASARPDQDCFL-----YGPAPAKTLP-----CPGPI 469
 QY 306 AGNFNLISGOCDDAOSPGILRLFOVLYOHPONESN-----KITV--VDLSNER 353
 DB 470 TOLHTFYVNGYNQOARTPG---YCIKIDH--ONSSINIESLELTITKISKNTDPCSK 523
 QY 354 AMSLTIEPPVYKSRKFEVGCFCVLESRCSSNL--TLNLSGSKHKSIFLDDLTRLMNY 411
 DB 524 NDSINDPR-LQPPKQFT--FL-LEQHSNEDQYKTLTT-EMHKT-----L 564
 QY 412 EKTISCTDHRVYCKRKSYSLOVPSDILHLPVELH---DESMKLLVPR-----DR 456
 DB 565 ELYNSSTTSEY--RKFTL-----IYHDESESVLFSVNPDIFFVEKFPVLVNSLKTDM 617
 QY 457 LS-----LVLPVAKLOQHTHEKPCNTSFLYASAIPODLYFGSCPGSGSIKOIO--- 508
 DB 618 LDMTGLSLSTIAHOKM-----NLTPAQVYVFA---NQAVKNQNL 656
 QY 509 -----YKONI-----SVTLRTFAPSPQOASRQGLVSFIP----- 539
 DB 657 RNMDLVKRQLEIFFTIADGVTTETALPQLELIRKMTNGRLIPGKTENSCTDNPPL 716
 QY 540 ---YFKEGVFTYT-----PDKSKYLLRTPMMDRGLPSLTVSMNLSVRDQ 584
 DB 717 IDGFEPEDGADFVQVGVGGLQNIKMQDSNGFVGNQENR----- 758
 QY 585 VACLTFFKERSGVVCGTGRAFMITOEORTAEELFSLDEDVLPKPSFHHSHFVNINSCS 644
 DB 759 ---LTF-----SGTIV-----TNSPEI-----AYKNPWF----- 779
 QY 645 PMSGKQLDLFSVTLPRVDTLVILIAVGGVLLSALGLIICVKKKKTKNGKPAV 704
 DB 780 -----ILFRIDSK-----LFAF---GVKLSALSTGGCQITVROKTSIGLV 819
 QY 705 G---LYNGNINTE-----MPR---QP---KKFOGR 726
 DB 820 GFTEINDDNVSTQIISORSSSDSMPIYIPMKYTSIGIIPKTPRLGHQPVNKKLAVAR 879
 QY 727 KDNDSHYAVIEDTMV-----YGHLLDSS----- 751
 DB 880 RBASSCSYNFISDSVVPKTELTTWTSASVGLVLRHIFHYOHLPADSSVCOGGMVD 939
 QY 752 ---GSFLQPEVDY-----YRPFQGTMGV--CP 774
 DB 940 KFCRCICPSKRYTGDYCMWRICHPPATYSFGICSCP 975

RESULT 9

VCVDAR
 new polyprotein - avian reticuloendotheliosis virus

N:Alternate names: coat polyprotein
 N:Contains: coat protein gp22; coat protein gp73

C:Species: avian reticuloendotheliosis virus
 C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 04-Dec-1994
 C:Accession: A03999

R:Wilhelmsen, K.C.; Eggleston, K.; Temin, H.M.
 J. Virol. 52, 172-182, 1984

A:Title: Nucleotide acid sequences of the oncogene v-rel in reticuloendotheliosis vir
 A:Reference number: A93003; MUID:85009850; PMID:6090694
 A:Accession: A03999

A:Molecule type: DNA

A:Accession: S57078
 A:Molecule type: DNA
 A:Residues: 1-818 <MAN>
 A:Cross-references: EMBL:449559; NID:g1015728; PIDN:CA89587.1; PID:g1015729; MIPS:YJR05
 R: Huang, M.E.; Manus, V.; Chuat, J.C.; Galibert, F.
 Yeast 12, 869-875, 1996
 A:Title: Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open reading frames
 A:Reference number: S71676; MUID:96437976; PMID:8840504
 A:Accession: S71681
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-818 <HUA>
 A:Cross-references: EMBL:L47993; NID:g1019675; PIDN:AA839285.1; PID:g1019681
 C:Genetics:
 A:Gene: SGD:PTK2; STK2
 A:Cross-references: SGD:S0003820; MIPS:YJR059w
 A:Map position: 10R
 C:Function:
 A:Description: involved in polyamine uptake
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
 C:Keywords: ATP; phosphotransferase; transmembrane protein
 F:253-553/Domain: protein kinase homology <Kin>
 F:467-483/Domain: transmembrane #status predicted <TMM>

Query Match 2.6%; Score 112.5; DB 2; Length 818;
 Best local similarity 19.2%; Pred. No. 5.3; Mismatches 317; Indels 281; Gaps 50;
 Matches 172; Conservative 127;

QY 100 MSGPCPEVQLOPSTSLPTLRTPTMDVKAH--KSIGLELOFSIPRLROIGPESCP 156
 1 MANGKDKKEVDKSPSVSTLKLGR-LFNSSHTDSSLLSAE-----QLNGRSLR 52
 DB 157 DGVTH-SIG-----RIDATVRA-----IGTFCNGTV-----SRIMQE 190
 53 KRPTSPSISGSGSGSSPSSAGARQSASLHRKKNASVG--FSNGSVSHKSSVALQD 110
 QY 191 GVKMALHLPMFH-PRNV--SGFSIARRSSIKRLCIESVEGEGSATLMSANYPEGPE- 246
 111 LKHN-NNYLYNSPDI LTGTGIASTRDR-AYLDREKEKERANKERHHAAGLPOR 168
 QY 247 -DELMTQF-----VPAHLRASVSF--LNFILNCERKEERYEY-----284
 169 SNSMASHRPENIYVNPYGISPNHARPTAFADTLNTN-----KENDLSFYMHGNSK 222
 QY 285 ----IPGSTNPEYFKLEKQPGMAGNFUSLQGCQDQASPGILRLQOVLYQHPONES 341
 223 IRMLPLRIANPNDFLPEDKQYSVHLTDNFVEDTDKPIGSGGSEVR-KVKSYSROKD- 280
 QY 342 NKIYVVDLSNERAMS/LTIPRPVKOSRKFPVCGFVLESRTGSSNLTLTSGSKKI- 397
 281 --VYAL-----KLNMTIYESPEKTYR-----CSKEFIIAKHLSHNVHTN 320
 QY 398 -----SFLCD-----DLTRL-----WMNVKTI SCTDHRVCO RKSYSIQ 431
 321 TFLVLTVPPTTYYTRGWMGFMELGVKDLQOLMERTGMKNP-----FNEKYCLFKQVAG 375
 QY 432 V-----PSDILHP--VELHDF--SWKLLVPKRLS-----458
 376 IFCHDNGIAHRDLKPEVNLISKEGICKLTDFGISDMYHIIPIHDYTS PVKTCQGMIGSP 435
 QY 459 -----LVLVPAOKLOOHTHEKPCN--TSFSY-----LVASATPSODLYFGSPFGGS 503
 436 YTPPEVWYFDKKNYPEKFOKRPYNPLAMDYSALGIMLTMINIIP-----FTDSC--- 486
 QY 504 IKQIQVKONISYTLRTFAPSFOEASROGLTWSFIRYFEEGVFTVTPDKSKVYLRTPN 563
 487 -----NTDARREFEVSYDNFINQN-----PHFRDKGCHKRPGSEYSLANFNK 532
 QY 564 WDRGLPSLTVSWNISVPR-----DOVACLTFEKERSGVVCGTGRAMITIOEORTAE 616
 533 TD-----ATRIAMRLADPMPATRYTMDLFDNDFFOQIEFCVPERNDLDLVRELR-KSP 586
 QY 617 EIFSLDEVDLPRKSPFHHSFWNISCSPTSGQDLDFSVTLGTRPTVDLTVLIAVGS 676

DB 587 STNDFSENSLDAP---HDEEVITSN--PELKKE-----TLTSKPRS-----MLEIAESP 631
 QY 677 GVLLSALGLITICVKKKKKKTKKGPAVGIYNGNINTEMPROPK-----FOGRKD- 728
 DB 632 SLKQKS-----FVKDSAKTKTTHDVGDEGENESTKPKQODKKENLKKDEYKNGDKDK 682
 QY 729 -----NDSHYVAYIE-----DFWYVGHLLQDSSGSFLQPEVDYRPFQGTMGVCPSPPT 778
 DB 683 VIEEATTTNVDSTLEKPTPTSTKVEDNLSDDS-----TMKELKSMNSTPTTP-- 731
 QY 779 ICSRAPT---AKLATEPPPRS-----PPESESPYTFSHPNNGDVSSKDTPIPLNT 828
 DB 732 -THNGPTPLPAKAGTQDKRMSDLKSETPASTKNPSAPN---VSSSSNLSLGS 784

Search completed: November 11, 2002, 11:33:54
 Job time : 60 secs

This Page Blank (uspio)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 11, 2002, 05:03:00 ; Search time 21 Seconds

(without alignments)
1651.153 Million cell updates/sec

Title: US-09-899-569A-4
Perfect score: 4394
Sequence: 1 MAGLNCGVSLALGVLLGA.....SSKQTDIPLNTQEPMEPAE 836

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_40:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	2.7	2499	MPRI_BOVIN	P08169 bos taurus
2	117	2.7	2476	ZAN_PIG	Q28983 sus scrofa
3	115.5	2.6	849	PER_AWPE	Q17062 antheraea p
4	115	2.6	582	ENV_AVIRE	P03399 avian retic
5	114	2.6	1328	YHV5_YEAST	P38851 saccharomyc
6	113	2.6	3329	BRC2_MOUSE	P97929 mus musculu
7	112.5	2.6	440	OMGP_HUMAN	P23515 homo sapien
8	112.5	2.6	818	PTK2_YEAST	P47116 saccharomyc
9	112.5	2.6	1805	RML_HUMAN	O92545 homo sapien
10	112	2.5	1694	SN_MOUSE	O62230 mus musculu
11	111	2.5	794	TRK1_LYMS	O76897 lymnaea sta
12	110.5	2.5	704	CRAR_MOUSE	P98064 mus musculu
13	110.5	2.5	942	TMK1_ARATH	P43398 arabidopsis
14	109.5	2.5	692	DPB2_YEAST	P24482 saccharomyc
15	108.5	2.5	1828	MAP2_MOUSE	P20357 mus musculu
16	108	2.5	699	CRAR_HUMAN	P48740 h complemen
17	108	2.5	838	SMAG_HUMAN	O98099 homo sapien
18	106	2.4	777	UNC8_CAELO	O21574 caenorhabd1
19	106	2.4	830	YBMA_SCHPO	O10332 schizosacch
20	106	2.4	1091	NCAL_CHICK	P13090 gallus gall
21	105.5	2.4	913	CADA_CHICK	P24503 gallus gall
22	105.5	2.4	4590	FATH_HUMAN	O14517 homo sapien
23	104.5	2.4	810	ILAR_MOUSE	P16382 mus musculu
24	104.5	2.4	1490	CRK7_HUMAN	O98944 homo sapien
25	104	2.4	449	PCOL_HUMAN	O15113 homo sapien
26	104	2.4	4036	RRLP_DUGBY	O66331 dugbe virus
27	103.5	2.4	727	CTCF_HUMAN	P08922 homo sapien
28	103	2.3	2347	KROS_HUMAN	P08922 homo sapien
29	103	2.3	2700	ZAN_HUMAN	O94493 homo sapien
30	103	2.3	2768	THRG_HUMAN	P01366 homo sapien
31	102.5	2.3	457	CD4_MOUSE	P06332 mus musculu
32	102.5	2.3	457	CD4_SAISC	Q29037 salmistr sci
33	102.5	2.3	1163	ITRL_MOUSE	P24063 mus musculu

ALIGNMENTS

```

RESULT 1
MPRI_BOVIN
ID      MPRI_BOVIN      STANDARD:      PRT: 2499 AA.
AC      P08169;
DT      01-AUG-1988 (Rel. 08, Created)
DI      01-OCT-1989 (Rel. 12, Last sequence update)
DI      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Cation-independent mannose-6-phosphate receptor precursor (CI Man-6-P
DE      receptor) (CI-MPR) (Insulin-like growth factor II receptor) (300 kDa
DE      mannose 6-phosphate receptor) (MPR 300) (MPR300).
GN      IGFR2 OR M6P.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_Taxid:9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-8115411; PubMed-2963004;
RT      Lobel P., Dahms N.M., Kornfeld S.;
RT      "Cloning and sequence analysis of the cation-independent mannose 6-
RT      phosphate receptor."
RT      J. Biol. Chem. 263:2563-2570(1988).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      Killian J.K.;
RT      Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 1039-2499 FROM N.A.
RX      MEDLINE-87175648; PubMed-2951738;
RT      Lobel P., Dahms N.M., Breitmeyer J., Chirgwin J.M., Kornfeld S.;
RT      "Cloning of the bovine 215-kDa cation-independent mannose 6-phosphate
RT      receptor."
RT      Proc. Natl. Acad. Sci. U.S.A. 84:2233-2237(1987).
CC      -1- FUNCTION: TRANSPORT OF PHOSPHORYLATED LYSOSOMAL ENZYMES FROM
CC      THE GOLGI COMPLEX AND THE CELL SURFACE TO LYSOSOMES. LYSOSOMAL
CC      ENZYMES BEARING PHOSPHOMANNOSYL RESIDUES BIND SPECIFICALLY TO
CC      MANNOSE-6-PHOSPHATE RECEPTORS IN THE GOLGI APPARATUS AND THE
CC      RESULTING RECEPTOR-LIGAND COMPLEX IS TRANSPORTED TO AN ACIDIC
CC      PRELYSOSOMAL COMPARTMENT WHERE THE LOW pH MEDIATES THE DISSOCIATION
CC      OF THE COMPLEX. THIS RECEPTOR ALSO BINDS INSULIN GROWTH FACTOR II.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein. Lysosomal.
CC      -1- DOMAIN: CONTAINS 15 REPEATING UNITS OF APPROXIMATIVELY 147 AA. THE
CC      MOST HIGHLY CONSERVED REGION WITHIN THE REPEAT CONSISTS OF A
CC      STRETCH OF 13 AA THAT CONTAINS CYSTEINES AT BOTH ENDS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: J03527; AAA30455.1; -.
CC      EMBL: AF342811; AAL23908.1; -.

```

```
DR PIR; A30788; A30788.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000479; CIMR.
DR InterPro; IPR000562; FN_Type_II.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00878; CIMR; 13.
DR PRINTS; PR00013; FNTYPEII.
DR PRODOM; PD000995; FN_Type_II; 1.
DR SMART; SM00059; FN2; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
KW Transmembrane; Transport; Glycoprotein; Repeat; Receptor; Lysosome;
KW Signal.
FT SIGNAL. 1 44
FT CHAIN 45 2499
FT DOMAIN 45 2313
FT TRANSMEM 2314 2336
FT DOMAIN 2337 2499
FT REPEAT 45 170
FT REPEAT 171 327
FT REPEAT 328 478
FT REPEAT 479 629
FT REPEAT 630 771
FT REPEAT 772 933
FT REPEAT 934 1089
FT REPEAT 1090 1229
FT REPEAT 1230 1373
FT REPEAT 1374 1518
FT REPEAT 1519 1658
FT REPEAT 1659 1807
FT REPEAT 1808 1999
FT REPEAT 2000 2137
FT REPEAT 2138 2290
FT DOMAIN 1909 1951
FT CARBOHYD 120 120
FT CARBOHYD 409 409
FT CARBOHYD 444 444
FT CARBOHYD 552 552
FT CARBOHYD 590 590
FT CARBOHYD 635 635
FT CARBOHYD 755 755
FT CARBOHYD 879 879
FT CARBOHYD 959 959
FT CARBOHYD 1030 1030
FT CARBOHYD 1173 1173
FT CARBOHYD 1255 1255
FT CARBOHYD 1321 1321
FT CARBOHYD 1665 1665
FT CARBOHYD 1766 1766
FT CARBOHYD 1825 1825
FT CARBOHYD 2094 2094
FT CARBOHYD 2145 2145
FT CARBOHYD 2220 2220
SQ SEQUENCE 2499 AA; 274526 MW; 3C1C9DEF2875159D CRC64;

Query Match 2.78; Score 118; DB 1; Length 2499;
Best Local Similarity 19.98; Pred. No. 3.2;
Matches 195; Conservative 100; Mismatches 343; Indels 340; Gaps 56;
```

```
Db 1863 AVCLISGSKGASRGRLA-----SMKLDYRHQDAVILSTANGCTCP 1903
Qy 219 -----RLCIIEYVEEGSATLMSANYPGFPEDELTQFVPAHLRASVSLFNLS 272
Db 1904 PETEGEPCVFPFVFENK-----SYECVYSRARLW-----CATNARD 1945
Qy 273 N----CERKEERYEYIIPGSTNPEVKLEDKQGNMAGNENLSLQCCDDAOSPILRL 328
Db 1946 HEMGFCKHSTSH-----RTSVILRK-----CDEAD----- 1971
Qy 329 QFYLVHPHONESMKIYVVDLSNERAMSLTIEPR-----PVKSRKFV----- 371
Db 1972 -----VRPQ-----VFSEVRGCEVFEMTKKYVCPKKKECFVQKHRTYDLRL 2017
Qy 372 ---PGCF-----VCLF-----SRCCSNLTL-----TSGSKRISF-----LC 401
Db 2018 SSLGSMSPFHNGASYYINLQKIKYKPODCEBASCKSTSEGVGLVHTQKLDVY 2077
Qy 402 DDLTRLMMANVEKTIISCDHRYCQ-----RKSYS-LOYPSDILHLPVLDHPSW 448
Db 2078 DD--RVILVYSKGYHCGDNKTASAVIELCAKTVGRSPSTRFVDSCSTYH-----FSW 2128
Qy 449 KLVPRKRLSLVLPAAKIQ--QHTHEKPCN-TSESYLVASAIKPSODLYGSPCGGSIK 505
Db 2129 -----DSRAACAVKPOEVQVNGVTTPNANGRSFSL-----GDIYKRFESAGDVR 2174
Qy 506 -----QIQVK-----ONISVTLRTFAPSFOEASRGTLVSPFIPYKEBG-- 545
Db 2175 TNGDRIYELIQLSITGSSPASCASICOR---KANDQHFHSRKVGTSNQRYYVQDGL 2231
Qy 546 --VFT-----VTPTKSKVYLK-TPNMDRGLPSLTSWMSNLSVRDQVACLTFFER 594
Db 2232 DVFTSSSKGCKDKTKSVSSTIFFHCDPLVKDGIPEFS-----HETADCYLPSWH 2282
Qy 595 SGVVCQGRAFMIQEQRTAEIFSLDEVDLPKPSHHSHFWNINCSPTSOKDILL 654
Db 2283 TSANCPGAGF-----DETAGDAGEHK-----GLERQAGVAVLSTL 2332
Qy 655 FSVTLPRPYDVLVILAAVGGVLLSALGLIIC-----VKKKKKTKNPGAVIYNG 709
Db 2323 ---LVALTCLLTLVLLKKERREVMRS--LTNCCRANAVSYKYSKVNEEBA----D 2372
Qy 710 NINTE-----MPQPKKFKQGRKNDSHYAV---IEDTM--VYGHLLDSSSFLQ-P 757
Db 2373 ENETEMLMEEIOPAPRPRGEGE--NGHVAAKSVRADTLSALHGC--EDDSEDEVLTLP 2429
Qy 758 EVDIYRPFQ--GTNGVCPSPPTICSRAPTK-----LATEPPRPPSPPESEPTFS 809
Db 2430 EVKVRPPGRAPGAGC--GPLRLPLPRKAPPLRADRVGLVRGEPARRGRPAAATPISTF 2488
Qy 810 HPNNGDVSSKDTDIPILN 827
Db 2489 H-----DDSDDELH 2498

RESULT 2
ZAN_PIG ZAN_PIG STANDARD; PRT; 2476 AA.
AC Q28983;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 823-830; 859-872; 883-890;
RP 920-925; 960-967; 1235-1244; 1349-1354; 1518-1532; 1624-1656;
RP 1658-1667; 1777-1795 AND 1914-1921.
RC STRAIN=Meishan; TISSUE=Testis;
```



```

GN PER.
OS Anthraea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Dityrisia; Bombycoidea; Saturniidae; Saturniinae; Saturniini;
OC Anthraea;
OX NCBI_TaxId=7119;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Head, and Thorax;
RX MEDLINE=95033232; PubMed=7946353;
RA Rappert S.M., Tsai T., Roca A.L., Sauman I.;
RT "Cloning of a structural and functional homolog of the circadian clock
  gene period from the giant silkworm Anthraea pernyi.";
RL Neuron 13:1167-1176(1994).
CC -1- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS. THE
  BIOLOGICAL CYCLE DEPENDS ON THE RHYTHMIC FORMATION AND NUCLEAR
  LOCALIZATION OF THE TIM-PER COMPLEX. LIGHT INDUCES THE DEGRADATION
  OF TIM, WHICH PROMOTES ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE
  HETERODIMER COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION
  NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN
  TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING
  INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY). EXPRESSION
  EXHIBITS PROMINENT CIRCADIAN VARIATION IN ADULT HEADS AND IN
  PARTICULAR IN THE PHOTORECEPTOR NOCLEI.
CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMELESS (TIM); THE COMPLEX THEN
  TRANSLATES INTO THE NUCLEUS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.
  INTERACTION WITH TIM IS REQUIRED FOR NUCLEAR LOCALIZATION (BY
  SIMILARITY).
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U12769; AAA64675.1; -.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS_domain.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS0112; PAS; 2.
KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation.
FT DOMAIN 63 74 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 155 225 PAS 1.
FT DOMAIN 300 370 PAS 2.
FT DOMAIN 378 418 PAC.
FT DOMAIN 53 58 POLY-SER.
FT DOMAIN 71 74 POLY-LYS.
SQ SSQUNCE 849 AA; 94799 MW; 07C8654EEC058770 CRC64;

Query Match 2.6%; Score 115.5; DB 1; Length 849;
Best Local Similarity 18.6%; Pred. No. 1;
Matches 167; Conservative 110; Mismatches 311; Indels 311; Gaps 42;

```

```

OY 228 EGGGSATLMSANYPEGP-----EDELMTQWVPA---- 258
OY 239 ENGSWTVCIRIRYRGISGFSVKNNTTAVPLAFKFNVEDKGNVLYIQAAPFE 298
OY 259 -----HLRASVFL-----NFNLSNCRKEERYEY--YIGGSTINPEVFLDEKOPG 303
OY 299 SAKTSNEVLAKTVPFIRHSADGNLEYID--AESVYLGYLPQDITRRALL--YHPG 354
OY 304 NMAGNFNLSLOGCDODASPGILRLQFVLPQHPQNSNKIYVVDLSNERAMSLTE--- 360
OY 355 DL-----GYLQETIGSLVKKGNTYRSTYTNMNGNMYVETEMSA 396
OY 361 -PPVQSKRKEVGCFCYLES-----RTCSSNLTLTSGSKRISFLCDLTRLMMNV- 411
OY 397 FIPWMSKLEFVYAGKHIIIGPANPDVFQNPENVLKLTEQKNQAKMYRDSIIRIMDV 456
OY 412 -----EKTISTDRHYCQKRSYLOV-----PSDILHLFVELHDFSWKLVPDR 457
OY 457 TKPAELAKQMS-----KRCODLAHFEMMLEEQPKVPDDLRLLEIQADHSY----YERD 507
OY 458 SLVLPVPAQKLOQTHHEKPCNTSFSYLAASAIPQDLYFGSFCGSGSIKQVKNISVT 517
OY 508 SVLGGISPHHEYDSKSTETPLSY-----NQLATYNDN-----L 541
OY 518 RTEAPSFQDASBQGLTVSEIPYFKEGVETVPDRKSKYLRTPMD---RGLPSLTV 574
OY 542 QRFNFSHQNSA-----FVDNML--PSRNPYLALAPHSSESIKNVPAMEY 585
OY 575 SWNIVPRDQVACLTFFKESGVY-----CQGRAMIIQEQTR--AEIPLSDE 623
OY 586 S-----GDVIDLTGEGTSGVIVFNKSPMGLKTKGPIRLTESSLTKHNAE---E 633
OY 624 DVLPRKPSFHHSF-----WNWISNCSPTSGKOLDLFSVTLTPRVDTVLILAAVGGVL 679
OY 634 KELMKIHKREKTSKGDRAVVSNEARQKKE-----HLARCNAGFQ 674
OY 680 LLSALGLIICVKKKKKKTKKGPAGIYNGNINTEMRQPKFKQKGRKNDSHYAVIED 739
OY 675 TISA-----ANNTPSV--YEKPHN--LKRSSKQMSSEPIAKNNH---CPS 712
OY 740 TMYVGHILDSSGSFLOP-----EVDTPYPPQGTMGVCPSP-----PT 778
OY 713 SROFRKQKTTCSGFAQPPSATNPVSTSSQSSSPVNNVNFILGVGMQPPMLPLSP 772
OY 779 ICSRAP-----TAKLATEPPRPSPESESEPTFSHPNNGVSSKQDTPILNTQEP 831
OY 773 VSGMFPMTYTPVTAITYTTSGRP-----SEPN--YHRNN--MNNNQFOQPLGNSRLP 820

RESULT 4
ENV AVIRE STANDARD; PRT: 582 AA.
ID ENV AVIRE
AC P03389;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide (Coat polypeptide) [Contains: Coat protein GP73; Coat
  protein GP22].
GN ENV.
OS Avian reticuloendotheliosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaetrovirus.
OX NCBI_TaxId=11636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A.
RX MEDLINE=85009850; PubMed=6090694;
RA Wilhelmisen K.C., Eggleston K., Temin H.M.;
RT "Nucleic acid sequences of the oncogene v-rel in
  reticuloendotheliosis virus strain T and its cellular homolog, the
  RT proto-oncogene c-rel.";
RL J. Virol. 52:172-182(1984).
CC -1- PTM: SPECIFIC ENZYMOLOGIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

```

CC -1- MISCELLANEOUS: STRAIN A IS A HELPER VIRUS OF THE STRAIN T.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: K02537; AAA9198.1; -
CC EMBL: X01455; CAA25686.1; -
CC PIR: A03999; VCDAR.
CC HSSP: P03385; 1MOF.
CC InterPro: IPR002050; Env_polyprotein.
CC Pfam: PF00429; Env_polyprotein; 1.
CC Coated protein; Glycoprotein; Transmembrane; Polyprotein.
CC CHAIN 1 391
CC FT CHAIN 392 582 COAT PROTEIN GP22.
CC FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 582 AA; 64138 MW; CD2560ADF026D32 CRC64;
Query Match
Best Local Similarity 2.6%; Score 115; DB 1; Length 582;
Matches 89; Conservative 60; Mismatches 147; Indels 170; Gaps 19;
OY 314 OGCDQDAQSPGLRLQFOVL-VQHPONESNKIVVDL-----SNERAMSLTEPRPVK 365
DB 221 RGVLDLPQTSIDLEATHQVLAATNPNQLAENCMCLTLGTSQPSRMRMSLSME----- 275
OY 366 QSKRPVPGCVLESSTCSNLTLTGSKKIS---FLC---DDLRLMKNVEKTSCTD 419
DB 276 -----IAVLASLSGATHRVNRCOLLCREADNRIGIPGVVHTNCTIS 317
OY 420 HRYCOKRSYSLOVPSD-----ILHPELCHDFSMKLLVPRDRSLVLPVQKQOHTHEK 474
DB 318 IQSSLRRLRYVELRDYVLRWVLYCEQNAVY---ALPKWKIGL----- 359
OY 475 PCMTSFSYLVASAIPODLYFGSCFPGSGSIKIQVKNONISVTLRTFAPSFOEASRQGLT 534
DB 360 -C-----ILASIVPDM-----SLTPG-----EERIPL-----PSLEYTAGNRKRA 393
OY 535 VSIFPFKEGEVFTV-----PPTKSK 556
DB 394 VQPIPLVGLGIGATLAGGTGLGVSVHTYHKLNSOLIEDVQALSGTINDLOQIDSLAE 453
OY 557 VYLRTNMDSGLPSLVSNINISVPRDQVACLFEFKERSGVGOTGRAFIIDQORTAE 616
DB 454 VVLQN---RRGLDLTLAEQGGICLAEQKCC-FYANKSGIV-----RD 492
OY 617 EIFSLEDDVLP-KPSFNHHSFWNINISCSPTSGKOLDLFSVTLTPRTVDLTVLLAAVG 675
DB 493 KIKKLQEDDLARKRALYDNLNMGNGFLPYLLPSLGRPL----- 532
OY 676 GGVLLISALGLIICVKKKKKTKNGPANGIYNGINTENPRQPKK 721
DB 533 -GILFLTLG---PCIRKTLTRIHHKIQGSKNPRISPAPQATPNR 574
RESULT 5
YHV5_YEAST
ID YHV5_YEAST STANDARD; PRT; 1228 AA.
AC P38851;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 143.6 kDa protein in SPO16-REC104 intergenic region.
GN YHR155W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RC MEDLINE=94378003; PubMed=8091229;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Doyer J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucada T., Hiller L., Jier M., Johnston L., Langston Y., Mouser L.,
RA Latteille P., Louis E.J., Macri C., Mardis E., Meneses S., Musher L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII"; 265:2077-2082(1994).
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: STRONG, TO YEAST SIP3. SOME, TO S. POMBE SPAC19A8.02.
CC -1- SIMILARITY: CONTRAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U01397; AAB68977.1; -
CC PIR: S46754; S46754.
CC SGI: S0001198; YHR155W.
CC InterPro: IPR001849; PH.
CC Pfam: PR00169; PH: 1.
CC SMART: SM00233; PH: 1.
CC PROSITE: PS00003; PH_DOMAIN; 1.
CC Hypothetical protein.
CC FT DOMAIN 308 421 PH.
CC SEQUENCE 1228 AA; 143583 MW; C8872EAE8270A4B6 CRC64;
Query Match
Best Local Similarity 2.6%; Score 114; DB 1; Length 1228;
Matches 99; Conservative 85; Mismatches 196; Indels 148; Gaps 21;
OY 179 SNGTVSRIRKQEVKALHFMFHPRVNVSFSTIANSSIKRLCTIS---VPEGGSAT 234
DB 301 SNGNSEEKLSG-----WLYMKTYYGHDKRRVYVVARMKFLQNNFVGSLSPSKT 352
OY 235 LMSANTPEGFPEDELMTQFVPAHLRAVSFLNLSNCRKEEREVEYIIPGSTNPEV 294
DB 353 YVEETDKFG-----ILMTIV-----EYLP----- 371
OY 295 FKLEDKQPGMAGNFNLSDGCDQDAQSPGLRLQFOVLVQHONESNKIVVDLSNERA 354
DB 372 -----KEPRNFCFKRLQNPNCNCTEENTYI-----DIIQAQSIDELKSWINTLTSHKR 421
OY 355 MSLLT-----EPRVQSRKRPVPGCFVLCESRTCSNLTLSGSKHISFLCDDLTRLMN 410
DB 422 IALSTKEENDPKYQLARKKIEPQFFEPASSSTIDKLITSFSKTLF-LVEELKKNYMS 480
OY 411 VEKTIQCTDHRVCOQRKSYSLQVPSDILHLPVELHDFSMKLLVPR----- 454
DB 481 EDDIYSLIDKKAHLRAVRSIPATQTLHAL-----PSTFLSVNGYPCATQATGWGANW 536
OY 455 DRLSLVLPAAKIQOHTHEKPCNTSFLVVAASIPS-----QDIYFGSFCPGSGSIKO 506
DB 537 NDLSYLVNPLKGSVH---KPAIVNSNSRFSVSYPDVPYSLKVDIDIOFSIFSIVNHR 593
OY 507 IOYKONISVTLR---TFAPSFQOE-ASROGLTVSFIPEKEGEVFTVTP-----DKRSKY 557
DB 594 LQVPKEL-VLLRVSSVWCPPNKKRFPASMAFVTLNHIYVYINISGFSYLRIDLDLDSIE 652
OY 558 YLRTP-----NMDRGLPSLTSSVSNISV-----PRDOVA 586
DB 653 YDKSPKVVSRMLHMQRG---DGLRPNMSVFTTDRRAVASKIQLIENKAMHPIRGEKE 708


```

Db 1295 KMTKHEDYSTSSGRNNLENSDSMS-----TSGPVYIHKGDSDL----- 1334
Qy 353 RAMSLIEPRPVKQSKKEVPGCFVLCLESTCCSNLTLTSGSKHKISFLCDLTRL-WMNV 411
Db 1335 -----PADQGSK-----CPECTQYAREENTO-IKENIS-----DLTCLIEIMKA 1372
Qy 412 EKTISTDHRVCOQRKSVSLQVPSDILHLVELHDFSM----- 448
Db 1373 EET-----CKMSSDKKQLPSPDKMEQNKKEFNISFQASGKNTRYESLKNVNIEN 1424
Qy 449 ----KILVPRKRL-SLVLPVPAOKLQOHT--HEKPCNTSFSYLVAASIPQDLYFGSGFCG 501
Db 1425 RENDELTLVSDLSNKLINGINKDKMHTSCHK-----AISKVVEEDHF-Pi 1471
Qy 502 GSIGQIQVQONISVTLRTAPSFQOFASNOGLTVSF-----IPYFKE--EGVFTVTPD 552
Db 1472 VVTSQLPAAQH-----PEYEIESTKEPTLLSFHTASGKKVKIMQESLDVKMLFDE 1522
Qy 553 TSKVYLR-TPMMDRLPLSLTYSNMNISVPRQVACLTFFPKERSGVVCOGRAFMIIQEQ 611
Db 1523 TO---YVRKTASFSGSKPL-----KDSKKELTIAVER-----IEVT 1556
Qy 612 RTRAEI---FSLDEVLPRKPSFHHHSFWNINSCSPTSQK-QLDLFSVTLTPRTVDLT 667
Db 1557 ASKCEMOMNFVSKETEMLPQONHMYROTENLKTSGNKGVENIENNENKPR----- 1611
Qy 668 VVILAAVGGVLLLSALGLITCCVKKKKKTKNGKPAVGIYNGINIMTEMPROPKFKQGRK 727
Db 1612 -----ICCIQSSYPVEDSALVTEDSRKTCAVE--SLSKGRK 1650
Qy 728 -----DMSHVAVVIEDTMYV-GHLLQDSSGSFLQPVDPVRYRPRQGMGCPPS 775
Db 1651 WLRQGDKLGTRNTITIECVKHEPDEFAGNASTESHVLIITREIDTNHVSQVSTL-IS 1709
Qy 776 PPTIC 780
Db 1710 DPNVC 1714

```

```

RT "Structure and chromosomal localization of the gene for the
RT oligodendrocyte-myelin glycoprotein."
RL J. Cell Biol. 111:2673-2679(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RN Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
RN [5]
RX MEDLINE-88198371; PubMed-3283151;
RA Mikol D.D., Stefansson K.;
RT "A phosphatidylinositol-linked peanut agglutinin-binding glycoprotein
RT in central nervous system myelin and on oligodendrocytes."
RL J. Cell Biol. 106:1273-1279(1988).
CC -1- FUNCTION: CELL ADHESION MOLECULE CONTRIBUTING TO THE INTERACTIVE
CC PROCESS REQUIRED FOR MYELINATION IN THE CENTRAL NERVOUS SYSTEM.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: OLIGODENDROCYTES AND MYELIN OF THE CENTRAL
CC NERVOUS SYSTEM.
CC -1- PTM: O-GLYCOSYLATED IN ITS SER/THR-RICH REPEAT DOMAIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 8 LEUCINE-RICH REPEATS (LRR).
CC -1- CAUTION: DO NOT CONFUSE OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG)
CC WITH MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M63623; AAA59970.1; ALT_SEQ.
CC EMBL; X57436; CAA40684.1; -.
CC EMBL; X51694; CAA35991.1; ALT_SEQ.
CC EMBL; BC018050; AAH18050.1; -.
CC PIR; A30187; A30187.
CC PIR; A34210; A34210.
CC PIR; A36688; A36688.
CC PIR; A39613; A39613.
CC Genew; HGNC:8135; OMG.
CC MIM: 164345; -.
CC InterPro; IPR001611; LRR.
CC InterPro; IPR000372; LRR_Nterm.
CC InterPro; IPR003592; LRR_out.
CC InterPro; IPR003591; LRR_tyr.
CC Pfam; PF00560; LRR; 6.
CC Pfam; PF01462; LRRNT; 1.
CC PRINTS; PR00019; LEURICHP.
CC SMART; SM00370; LRR; 1.
CC SMART; SM00013; LRRNT; 1.
CC SMART; SM00369; LRR_tyr; 2.
CC KEGG glycoprotein; Myelin; Cell adhesion; Repeat; Leucine-rich repeat;
CC signal; GPI-anchor.
CC SIGNAL 1 24
CC CHAIN 25 440
CC REPEAT 33 54
CC REPEAT 55 77
CC REPEAT 78 100
CC REPEAT 122 145
CC REPEAT 146 168
CC REPEAT 170 189
CC REPEAT 190 213
CC REPEAT 217 228
CC REPEAT 229 270
CC REPEAT 271 292
CC REPEAT 293 335
CC REPEAT 336 377
CC REPEAT 378 416
CC CARBOHYD 45 46
CC CARBOHYD 61 61
CC CARBOHYD 103 103

```

```

FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 440 AA; 49608 MW; 762FE08905EA050 CRC64;

Query Match
Best Local Similarity 19.1%; Pred. No. 0.66;
Matches 79; Conservative 62; Mismatches 144; Indels 129; Gaps 17;

QY 256 VPAHLRASVSFLNF-----NLSNCKRERRE--YYIPGSTNPVEV-- 294
D 49 LPSGLQENIIHLNLSYNHFTDLHNLQYTLRLTLDISNNLLESLPAHLPRSLMNMGAAN 108
QY 295 --FPLEDKQPGNMGNNLSLQGDQD-----AQSGLRLQFOVLYVHPONES 341
D 109 NNILDK--SDYAYQNNLKYLDVSKNNLEKVLKNTLRLEVLNLSNNKLTVPPTNMP 166
QY 342 NKIYVVDLSNERAMSLTEPRPVKSRFVGCFCLESRTCSNLTLSGSKHKISFLC 401
D 167 SKLHIVDSLNN---SLT-----QILPGTLI-----NLT----- 191
QY 402 DDLTRLMMANVEKTSICDHRVQKRSYSLOVPSDILHLPVELHDFSKLLVPKDRLSVL 461
D 192 -NLTHLLHNNKFFTFIDQSFQ--LFQDQ-----ELTLYNNRMS--CDHKQNTIYLL 239
QY 462 VPAKLGQHTHEKPCNTSEFYLVA-SALPSSDLAFGSPCGSGIKQOVONISVTLR-- 518
D 240 KMMETKAHVIIGTSPQISLKEHNNYPPPSGTSLSLFTVSGQYDTJNISLVVQPK 299
QY 519 -----TFAPSFQOESRQGLTVEFIY-----FKEE 544
D 300 VTKIPKQYRKRETFEGATLSKDTFTSTDKAFVYPEDTSTETINSHMAAATLTILQD 359
QY 545 GVFT----VTPTKSKVYLRTPNDRGLPSLTYSWNISVPRQVACLTFEKKERS 595
D 360 GMYNTSLTSLSTKSSPTMPLTSLTSGMPN-----NFSEMPQOSTTILNMEET 407

RESULT 8
PRK2_YEAST STANDARD; PRT; 818 AA.
AC P47116;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase PRK2/STK2 (EC 2.7.1.-).
GN PRK2 OR STK2 OR YJR059W OR J1725.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=97079202; PubMed=8920934;
RA Nozaki T., Nishimura K., Michael A.J., Maruyama T., Kakimura Y.,
RA Igarashi K.;
RT "A second gene encoding a putative serine/threonine protein kinase
RT which enhances spermidine uptake in Saccharomyces cerevisiae.";
RL Biochem. Biophys. Res. Commun. 228:452-458(1996).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=96457976; PubMed=8840504;
RA Huang M.-E., Manus V., Chuat J.-C., Gallibert F.;
RT "Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open
RT reading frames and a gene cluster with a counterpart on chromosome
RT XI.";
RL Yeast 12:869-875(1996).

```

```

RN [3]
RP CHARACTERIZATION.
RX MEDLINE=97299648; PubMed=9154797;
RA Kaouass M., Audette M., Ramotar D., Verma S., de Montigny D.,
RA Gamache I., Torossian K., Poulin R.;
RT "The STK2 gene, which encodes a putative Ser/Thr protein kinase, is
RT required for high-affinity spermidine transport in Saccharomyces
RT cerevisiae.";
RL Mol. Cell. Biol. 17:2994-3004(1997).
CC -1- FUNCTION: ESSENTIAL DETERMINANT FOR HIGH-AFFINITY SPERMIDINE
CC TRANSPORT.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D87274; BA13325.1; -
CC EMBL; Z49559; CAAB9587.1; -
CC EMBL; I47993; AAB39285.1; -
CC SCD; S0003820; PRK2.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC Prodom: PD000001; Euk_pkinase; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
CC DOMAIN 255 562
CC NP_BIND 261 269 ATP (BY SIMILARITY).
CC BINDING 285 285 ATP (BY SIMILARITY).
CC ACT_SITE 388 388 BY SIMILARITY.
SQ SEQUENCE 818 AA; 91400 MW; 55BC8682894FB4BD CRC64;

Query Match
Best Local Similarity 19.2%; Pred. No. 1.6;
Matches 172; Conservative 127; Mismatches 317; Indels 281; Gaps 50;

QY 100 MSGCPFGYQLOPSTSLPLTNFTWYKAR---KSLGLELQFSIPRLQIGPESCP 156
D 1 MAGNCKDEVDKSPSVTLKLGKR-LFNSSHTDNLSSLSAE-----QLNGRSLR 52
QY 157 DGVTN-SISG-----RIDATVVR-----IGTFCNGTV--SRIMOE 190
D 53 KRPLSPSTISGSGSGNPSAGARQRSASLHRKKNASVG--FSNGSVSSHSVALQD 110
QY 191 GYKMAHLPMWFH-PRNV-SGFSIANRSSIKRLCIIEVEGEGSATIMSANYPEGPE- 246
D 111 LKINN-NNPPLNSDGLGTGTSIRDRD-AVLDRKEKEKARAKERTHTHAGLPOR 168
QY 247 -DELMTQF-----VPAHLRASVSF--LNFNLSNCKRERREYI----- 284
D 169 SNSMASHHFPENIVNYPGDISPNHARPDFAFDLTNTN-----KENDLSFYHMDNSK 222
QY 285 ---IPGSTPEVFKLEDKQPGNMGNNLSLQGDQDQAQSGTLRLQFOVLYVHPONES 341
D 223 IRMLPRLPIANPNLDPDMQYVHLTDNFVDTDNKPISSGGSSEVR-KVKSYSRKD- 280
QY 342 NKIYVVDLSNERAMSLTEPRPVKSRFVGCFCLESRTCSNLTLSGSKHKI---- 397
D 281 --YVAL-----KLNKTIYHSPKFTYR-----CSKEFLIAHLSNVAITN 320
QY 398 -----SFLCD---DLTRL-----WMNVEKTSICDHRVQKRSYSQ 431
D 321 TFYLKLVPTTTTTRGWFIMELGVKDLFQLMERTGKKNVP-----FNEKYCJLFQYAAQ 375
QY 432 V-----PSDILHP---VELHDF---SKLLVPPDRLS----- 458

```



```
Dy      376 IKCFHNDNIARDLKPEENVLLSKSEICKLTPDGSIDWVHVHPDVTSVKNGQMGISBP 435
OY      459 -----LVLPVPKKLOAQHTHEPCN--TFSFST-----LVA5AIPSDDLLFGSFCPGG 503
          ||| :||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      436 YTPREVWFADAKKHYPKEFKORYNPFLANDSYALGIMLTMINIIP-----FIDSC---- 486
OY      504 IKOIQOVKNISVTLTRTFAPSPFOGEASRGGLTVSFIPYKREGVFTVPTDTKSXYVLTRPN 563
Db      487 -----NTDAREFEFEEVSTDFNFTHNN-----PHERDKGCCHKPGGSEYSILARNFN 532
OY      564 WDRGLPSLTSYSWNISYVR-----DOVACLTEFERKESGVYCQTGRAFMIIOEORTBAE 616
          ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Cc      533 TD-----ATRLAMRLADNPATRMTMDLEFDPFEQQIETCYEPRDDDLVRRPELR-KST 586
OY      617 EIFSLDEVDLPKPFSEHHHSFWVNISNCSPSTSGQLDLLFVSYLTRPRTYLDVILLAAVG 676
Db      587 STPDFSENSLDAP-----HQDEVITHSN--PLFKKE-----TLTSKRPS-----MLEIASP 631
OY      677 GVLLLLSAGLIICCKKKKKKTNGRPAVGINGNTINTEMPQPKR-----EQGRGD- 728
Db      632 SLKOKS-----KVKDSAKTKTHDGDEGNSTSRPKODKKENLKDEVKKGDDXK 682
OY      729 -----NDSHVAVIE-----DTMYGHLLDSSGSFFLPEDVTYRPFGTGWCVCPPSPPT 778
          ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      683 VIEEATTNVDSILEKPFPRTSTKYEDNISBEDS-----TKMELKSMLSITPTTP-- 731
OY      779 ICSRAPT---AKLAETEPPPRS---PESESEPPTYFHPPNGDVDSSKTDPIPLNT 828
Db      732 -THNGPFTLPKAKAGTQLDKRMSDLSLKSETPASTKNFSAPN---VSSSSNSLSRLSGS 784

RESULT 9
RW1_HUMAN STANDARD; PRT; 1805 AA.
AC   Q92545;
DT   16-OCT-2001 (Rel. 40, Created)
DI   16-OCT-2001 (Rel. 40, Last sequence update)
DY   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Rml protein (Fragment).
GN   RML OR KIAA0257.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX   Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RX   NCBI_TaxID=9606;
RN   [1]
RP   TISSUE=Bone marrow;
RC   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohita M., Kawarabayashi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc   or send an email to license@lsb-sib.ch).
CZ   EMBL: D87446; BAAJ3387.1; -.
FT   NON_TER 1
SQ   SEQUENCE 1805 AA; 197588 MW; 955589214413AB84 CRC64;
```

Dh	485	SNLFAIINSP	PIELAIKSHIIIGDGI	SLIELVAENG	NGNTIIISL	PEPEKSSLD	QSSV	544											
Qy	85	-PENNR-----	-----	-----	-----	-----	-----												
Dh	545	TLASGYFAVE	RVKRLTAKLEG	IHDAIQI	ITWDEYELL	IPVAKAVI	ANGSLTCCPKKRVLP	604											
Qy	120	-----	TLN--RFTIMVKAH--	KSLGLEQF	SIPLRQIG	GEOSCPD	CVHISGR	166											
Dh	605	SFPGKIVHQ	SINIMNSPQKVI	IQQIRSL	SESDVRFYKRLR--	GKNEDE	PEPKSKIANI	662											
Qy	167	IDATVVRIG	FECSNG-----	-TVSRIKMQ	QGVKALHT---	PM-FHPR	NSGFSIARRS	218											
Dh	663	YFDPGLOCG	HCYGLGFLPS	KESEPKYQ	PGVAMQ	EDMDAD	MDLHOSLFGW	GKIGENSGH	722										
Qy	219	RLCIIESVE	FEQEGS	ATLMSAN	DEGPEDE	ELMTQFV--	PAHLRASV	FLNLSNCR	276										
Dh	723	RL-----	SAIFEVN---	TDLOKNI	ISKITAE--	LSMPSIL	SPRHLKEPILT-----	NTNCS	769										
Qy	277	KEE-----	RVETIY	IPGSTN	PEVYKLED	KDQPG	MAGNFNLS	IQGCDQDAOSGI	325										
Dh	770	EEETLLEN	PADVPYVQ	FIPAL	YSNPSEF--	VDK-----	IVSRFNLS--	-----	KVAKIDL	817									
Qy	326	LRLOFOVL--	-VOHQNS	NKIVYV	DLSENER	AMSLTEPR	PRVQOSR--	KVPGCFVCL	ESRT	382									
Dh	818	RTLEQYVFR	NSAHL--	OSSGTF	MEGSLR	HLILLLK	PEKKS	VAKKFLP-----	VHNRT	871									
Qy	383	CS-	-----	NLT-----	TSQSKH	ISFLCD	LTRLMM	NEKTI--	SCDHR	424									
Dh	872	VSSLIVRN	NNTVMD	AVMVGQ	GTTEL	RYAGKL	PEGSSL	RFKITE	ALNKC	TD-----	926								
Qy	425	RKTSLOV	PSDILH	VELH	DFSKLL	LVPKRL	SLVLP	QAOK	IQO	THEKPC	NSIFS	SYLV	484						
Dh	927	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	959						
Qy	485	ASAI	PSODLV	FGSP	CGSGI	KIQ	IOVRO--	NSVT-----	-LRT	APSRQ	QASHQ	GLTSFI	538						
Dh	960	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1006						
Qy	539	PYFKEGV	FTVTP	PTKSKV	-----	-LRPN	MDRGL	PLST	SVSW	INISV	PRD	VLATPEFK	592						
Dh	1007	TTTSS	EFILMAS	LPHML	ATCAEL	PRNWE-----	-----	-----	-----	-----	-----	-----	1044						
Qy	593	ERSGV	COTGRA	MI	IOE	QRT	RAE	IFSL	DE	DVLP	KPSP	HHSHF	VNINS	CNS	PTSG	KOLD	652		
Dh	1045	IISIM-	SALE	FLIV	IGTAV	LEAQ	GIW-----	-EPFR	RRL	SF--	EAS	NPF	DVGR	FPD	1092				
Qy	653	LLFV	TLPR	IVD	TVIL-----	-IANG	GVLL	LSA	GLIC--	-CYK	KKK	KTK	NGP	702					
Dh	1093	LRRIT	GISS	SEG--	NINT	LSIC	DPGHS	RGFC	GAGG	SSRPS	SGSHK	QCP	SVHP	SHS	NNNS	1151			
Qy	703	A--	-VG	ING	NINTE	MP	RQPK	FOG	KRDN	DSHY	AVATE	PTMY	GHLL-----	QDSG	752				
Dh	1152	ADVEN	RAKNS	SSST	SSTSA	QAAS	SGS	AKNT	IS	PLV	LDNST	VYQ	GH	AGK	SKGA	QSOHG	1211		
Qy	753	SFLQ	PEVD	YTR	PFQ	GTG	VC	PS	PTIC	SA	PTAK	LA	TEE	PR	SP	SESE	PTYFS	HPN	812
Dh	1212	S-----	QHHA	SPLE--	-QH	P	PL	PPV-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1252
Qy	813	NGD-----	-YSK	TDI-----	-----	PLN--	-TQ	ERME	834										
Dh	1253	HPER	ASSAR	HSE	DSIT	SLIE	AMDK	DFD	HHDS	PA	LEV	TE	Q	PP	1298				


```

FT CONFLICT 1065 1065 S -> O (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 1694 AA; 183087 MW; 1D2BFB3B7296C8C CRC64;

Query Match
Best Local Similarity 20.6%; Score 112; DB 1; Length 1694;
Matches 109; Conservative 64; Mismatches 191; Indels 166; Gaps 25;

OY 230 EGSATLSMANYDEGPEDLMTW-----QFVPAHLR-ASVFLNFMNSCKEEREY 281
DB EQQAVVLSGVPTGVSEGTYSYMYODGRPLQESTSTLIALISLQAGAYHCQAQAPDT 962
OY 282 EY-----YIPGSTNPEVFKLEKQPGNAGNFN-----LS 312
DB 963 ATASLAAPVSLHVSYPHRAVITLSALLSTPERIGHLYCVSQSPAPQQLQFHNRLVAST 1022
OY 313 LQCCDDAOS-----PGLRLQFOVLVQHPQESKRYIVVDLSN----- 351
DB 1023 LQGADELASGNRLHVTVPNELRLQ-----IHPELEDDGTVCASNTLQASAAAPD 1078
OY 352 ERAMSLTIPRPVKOSRKRFVPCFVLESRTCSNLTLSGSKHKRSFLCDDLTRM--- 408
DB 1079 AQAVRTVPMATVQEGQV-----NLTCLVWSTHQ-----DSLSTWYKG 1119
OY 409 -----MNVEKTISTCDHRYC-----QRKSYSLQVPSDILHLELHDFSMKL 451
DB 1120 GQQLGARSTILPSYKVLATATSYR-CGVGLPGHAPHLSPVTLIDVLAHARNLR-LTYLLE 1177
OY 452 VKRDLSTLVV-----PAQKIQOHTHEKPCMTSFSYVAS---AIP----- 489
DB 1178 TQGRQIATVLCVDSRRPQQLTSLHGDQ-----LVASSTEASVPMTLRLELDPPR 1228
OY 490 -GODLY-FGSPFCGS-----IKQIQVKNISVLTLPFAPSFOEASRQGLTSPFIP 539
DB 1229 SNEGILSCASHSPGLNANTSLLELLEGVAKNMPGSGSVEGEVTVTCEDPALLSALYA 1288
OY 540 YFK-----VEFTPDTKSKVYLRTPMNDRGPSLTSYSMNI-SVPRDQVACL 588
DB 1289 WFNHNGHWLOEGPASSLOFLVTRAHGALFCQVHDTQGRSSPAPLQTLIYAPRD--AYL 1346
OY 589 TFFKERSGVVCGTGAEMTIQ---EORTAEIIFSLDEVDLPKPSFHNS 635
DB 1347 SSFRD-----SKTRLVAVIQCTVDEPPEAEVLSHGKVLAAHSRHSS 1390

RESULT 11
TRKL_LYMWST STANDARD; PRT; 794 AA.
AC 076997;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 15-JUN-2002 (Rel. 41, Last sequence update)
DE Putative neurotrophin receptor LTRK 1 precursor (EC 2.7.1.112).
OS Lymnaea stagnalis (Great pond snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC Lymnaeidae; Lymnaea.
OX NCBI_TaxID=6523;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98232499; PubMed=95644036;
RA van Keesteren R.E., Fainzilber M., Hauser G., van Minnen J.,
RA Vreugdenhil E., Smit A.B., Ibanez C.F., Gereerts W.P.M.,
RA Bulloch A.G.M.;
RT "Early evolutionary origin of the neurotrophin receptor family.";
RL EMBL J. 17:2534-2542(1998).
CC -1- FUNCTION: MAY BIND AN ENDOGENOUS INVERTEBRATE NEUROTROPHIN. BINDS
CC HUMAN NT-3, BUT NOT NGF OR BDNF.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSION IS CONFINED TO THE CENTRAL NERVOUS
CC SYSTEM AND ITS ASSOCIATED ENDOCRINE TISSUES.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY.

```

```

CC -1- SIMILARITY: CONTAINS 2 LEUCINE-RICH REPEATS (LRR).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC EMBL; U61728; AAC26840.1; -.
CC HSSP; P08631; 1AD5.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001611; LRR.
CC InterPro: IPR000372; LRR_Nterm.
CC InterPro: IPR003591; LRR_Typ.
CC InterPro: IPR002011; RtkinaseII.
CC InterPro: IPR001245; Tyr_pkinase.
CC pfam: PF00069; pkinase; 1.
CC pfam: PF00560; LRR; 1.
CC PRINTS: PR00109; TYRKINASE.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART; SM00013; LRRNT; 1.
CC SMART; SM00369; LRR_Typ; 1.
CC SMART; SM00219; TYRK; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
CC Transferrase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
CC Phosphorylation; Receptor; Glycoprotein; Neurogenesis;
CC Leucine-rich repeat; Repeat; Signal.
CC SIGNAL 1 33
CC CHAIN 34 794
CC DOMAIN 34 419
CC TRANSMEM 420 440
CC DOMAIN 441 794
CC REPEAT 181 204
CC REPEAT 205 228
CC DOMAIN 504 775
CC NP_BIND 510 518
CC BINDING 538 538
CC ACT_SITE 647 647
CC CARBOHYD 64 64
CC CARBOHYD 102 102
CC CARBOHYD 128 128
CC CARBOHYD 288 288
CC CARBOHYD 374 374
CC MOD_RES 673 673
CC MOD_RES 677 677
CC MOD_RES 678 678
CC MOD_RES 789 789
CC SEQUENCE 794 AA; 89054 MW; FFF3EF576E1A440 CRC64;

Query Match
Best Local Similarity 20.7%; Score 111; DB 1; Length 794;
Matches 119; Conservative 62; Mismatches 182; Indels 212; Gaps 32;

OY 203 PRNVGSFIANRSSIKRLIIESVFEGEGSATLSMANYDEGPEDLMTWQFVPAHL-- 260
DB 4 PRFRRLMTRANVLTV--ISILTSILSGAGCSPLSQ-----LPSDN-----PAHVCV 47
OY 261 -----RASVFLNFMNSCKEEREVEYIIPGSTNPEVFKLEKQPGN 304
DB 48 ODGVTERVDRSKNHNRTTASSGAHRTVSGEPGLDGV---TTRSTVAP-----DQVPG 98
OY 305 -----MAG-----NPNLSLQGGDDAQ-----SPGILRLQFOVLVQHPQNE 340
DB 99 ASRNTTMAGKCSLOYDLSTFACPACQCNANISGVAVSCVPTDLR-EPVYARAVARA 157
OY 341 SNRIYVVDLSNERAMSLTIEPRPVKOSRKRFVPCFVLESRT---CSSN-----LTLT 390
DB 158 -----YIKLELRQOSKLT-----SLKTEIKF-----FICLHILTIENGCLNIGSIARKTL 204

```

QY 391 S-----GSHKISFLCD-----DLRLMNVK-TISCIDHRCQKRSYSLAVPSDILH 438
CC 1
CC 205 SLFTINLRNHNHLEFPQELLRLTINLRMLTFCSTIN-----LMLRSVDVAADRS 259
DB 439 LPELHNFMSKLVLPKDRSLVLPKADQLQOHHEKRCN-TFSYLVASAIPLSDLYFGS 497
QY 260 MCGSTRG-----VSKMKMVFQFCE-PCGIPDIRNMMLVFEPRKMFILR 303
DB 498 FCPGSGIK-OIQVKONISVTLRTFAPSFQOEASRGLTSGIFPKFE-----GVFTV 549
DB 304 FVTSCKPKRIDLRNHNHVRSSSQF-----LDFKSEFNGQVYTGITTI 351
QY 550 TP--DFKSKVYLRTPMWDRGLPSLTYSWNI SVPRQVACITFEKERSGVVCOGTGRAFI 607
DB 352 LPHMETSGTYY-----LTAVN----- 368
QY 608 IQORRAEIEFSLDEVDLPKPSFHHSFVNISNCSPTSGKOLDLFSVTLPR--TV 664
DB 369 ---SKGANQTFHLVDQTPASSIH-----IPLSNIPR-----ISSATTPRASPT 412
QY 665 DL---TVILAAVGGVLLSALGLITCYKKKK 696
DB 413 DFGPQTVIIPVGVVILLISAVFIILCQRAKHR 447
RESULT 12
CRAR_MOUSE STANDARD: PRT: 704 AA.
ID CRAR_MOUSE
AC P98054;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement-activating component of Ra-reactive factor precursor
DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (harp)
DE (Mannan-binding lectin serine protease 1).
GN MASP1 OR CRAR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE-Liver;
RA MEDLINE=94179811; PubMed=8133044;
RA Takayama Y., Takada F., Takahashi A., Kawakami M.;
RT "A 100-kDa protein in the C4-activating component of Ra-reactive
RT factor is a new serine protease having module organization similar to
RT C1r and C1s".
RL J. Immunol. 152:2308-2316(1994).
RN [2]
RP SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE-Liver;
RA MEDLINE=93176166; PubMed=8493319;
RA Takahashi A., Takayama Y., Hattuse H., Kawakami M.;
RT "Presence of a serine protease in the complement-activating component
RT of the complement-dependent bactericidal factor, Harp, in mouse
RT serum".
RL Blochm. Biophys. Res. Commun. 190:681-687(1993).
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR HARP
CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CC CASCADE BY CLEAVING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1- SUBUNIT: HARP CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
CC (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)
CC LINKED BY A DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: LIVER.
CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHT (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16492; BAA03944.1; -.
CC HSRP: P00736; IAPQ.
CC MEROPS: S01.198; -.
CC MGD: MGI:88492; Masp1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001254; Ser_protease TRY.
DR Pfam: PF00084; susht. 2.
DR Pfam: PF00089; trypsin. 1.
DR Pfam: PF00431; CUB. 2.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00032; CCP. 2.
DR SMART: SM00042; CUB. 2.
DR SMART: SM00179; EGF_CA. 1.
DR SMART: SM00020; Tryp_Spc. 1.
DR PROSITE: PS00010; ASX_HYDROXYL. 1.
DR PROSITE: PS01180; CUB. 2.
DR PROSITE: PS01186; EGF_2. 1.
DR PROSITE: PS01187; EGF_CA. 1.
DR PROSITE: PS02040; TRYPSIN_DOM. 1.
DR PROSITE: PS00134; TRYPSIN_HTS. 1.
DR PROSITE: PS00135; TRYPSIN_SER. 1.
KW Hydrolyase; Complement pathway; Serine protease; Protease;
KW Glycoprotein; Susht; Repeat; Signal; EGF-like domain; Hydroxylation.
FT CHAIN 1 25 704
FT FT
FT CHAIN 25 453
FT CHAIN 454 704
FT DOMAIN 25 143
FT DOMAIN 144 187
FT DOMAIN 190 302
FT DOMAIN 305 368
FT DOMAIN 371 438
FT DOMAIN 454 704
FT ACT_SITE 495 495
FT ACT_SITE 557 557
FT ACT_SITE 651 651
FT MOD_RES 164 164
FT DISULFID 78 96
FT DISULFID 148 162
FT DISULFID 158 171
FT DISULFID 173 186
FT DISULFID 190 217
FT DISULFID 247 265
FT DISULFID 306 354
FT DISULFID 334 367
FT DISULFID 372 419
FT DISULFID 402 437
FT DISULFID 441 577
FT DISULFID 619 636
FT DISULFID 647 677
FT CARBOHYD 34 34
FT CARBOHYD 183 183
FT CARBOHYD 390 390
FT CARBOHYD 412 412
SO SEQUENCE 704 AA; 79895 MW; 71F4F3012D2C67F CRC64;
Query Match 2.5%; Score 110.5; DB 1; Length 704;

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 942 AA; 102387 MW; 93E30B52FE549DE CRC64;

Query Match 2.5%; Score 110.5; DB 1; Length 942;
 Best Local Similarity 19.4%; Pred. No. 2.8; Mismatches 216; Indels 233; Gaps 29;
 Matches 128; Conservative 82;

117 LLPLNFTFMDVKAHKSIGLEOFISIPRLKQIPGSCPDGVNHSI-SGRIDATVVRIG 175
 16 LLLSLSA---DSDGDSAMLSLKSISPPSSFEWSPDPCKWTHIYCTGKRTYRQIG 72
 QY 176 TFCGNGSVRIKQEGVKALHLFWFHPRVNSG--FSIANRSTIKRLCTIESVEG---- 229
 Db 73 HSGLOGTSLDRLNLSLELRLELQW---NNISGVPVSLGSLAQVYMLSNNDSPSD 129
 QY 230 --ESASATLMSANPEGEPEDELMWQFVPAHLRASVSFLNFNLNCKEEREVEYIIPG 287
 Db 130 VFQGLTSLQSE---IDNNPFKSMW---IPESLNMAALNFSANSA----NYSGLPG 178
 QY 288 STNPEVFKLEDKOPG-----NAGNFNLISLOGCDDAQSPGILRLQFVLYQHPO 338
 Db 179 -----FLGPDPFGSLIHLAFNNLEGEIPLSLAS-----QVQSLMLNGQ 219
 QY 339 NESKRIYVDSLNERASLTLEPRPVKOSKRFVGCFCLESRTCSNLTLTSGSKHKIS 398
 Db 220 KLTGDIYV--LQNMWGLK-----EVLHNSKFGPLDPFSGLKL-- 257
 QY 399 FLCDLRLRMNWKITSCDHRCCQKRSYSLOYPSDILH-----FVEL 443
 Db 258 -----ESLSLRDN-----SFTGPVPSLLESLSKVNLTNNHLOGEPV 297
 QY 444 HDFSMLLVPRKDR-----LSLVLY-----PAKQLQOHTHEKPCNTSESY 482
 Db 298 FKSSVSYDLKDSNFSCLSPCEDPRVKSLILLASSFDPRLAESMKGDPTN---- 353
 QY 483 LVASAIQSQDIYFSGSPGSGIKOIQVKONISVTLRTFAPSFOQASRQGLTV-----S 536
 Db 354 -----WIGIACSNGNITVYISL-EKMELT-GTISEFGAIKSLQRIILGINMLTG 400
 QY 537 FIPFKEGVFTVYPRK-----SKVYLRPNMDRGLPLSTVSWNISVPRQVACLTF 590
 Db 401 MTPQ---ELTLPNKTLDVSNKLEGVKPGF----- 429
 QY 591 FKERSGVVCTGAFMIIOEQRTRAEIIFSLDEVDLPRKPSHHHSFWNINSCSPSG-- 648
 Db 430 ---RSNVV-----NTNGNPDIGKDKSLSP-----GSSPSGSG 463
 QY 649 -----KQDLLESVTLTPRTVDLTVLLIAVGGVLLSALGLIICCVKKKKKTKNG 701
 Db 464 SGINGDKDRGMKSTFEG-----IIVGSVLGLLSTIFLIGLIVPCWYKROKRRFSG 515

RESULT 14
 DPB2_YEAST STANDARD; PRT; 692 AA.
 ID DPB2_YEAST
 AC P24482; Q06622;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase epsilon subunit B (EC 2.7.7.7) (DNA polymerase II subunit B).
 GN DPB2 OR YPR175W OR P9705.7.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH8;
 RX MEDLINE=91271241; PubMed=2052544;
 RA Ataki H., Hamatake R.K., Johnston L.H., Sugino A.;

RT "DPB2, the gene encoding DNA polymerase II subunit B, is required for
 RT chromosome replication in Saccharomyces cerevisiae."
 RT Proc. Natl. Acad. Sci. U.S.A. 88:4601-4605(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgorge W.,
 RA Ataulio R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Brastein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo J., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hedling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunkle-Smith S., Hyman R., Johnston M., Kalmann S., Kleine K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Maerke R., Messenguy F., Mewes H.-W., Mitsuhashi S., Mostl D.,
 RA Mueller-Auer S., Nemach A., Newkirk U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tetteh H.,
 RA Urestarez L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hanl J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
 RL Nature 387:103-105(1997).
 CC -1- FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
 CC REPLICATION. DPB2 IS ESSENTIAL FOR CELL GROWTH. MAY HAVE A ROLE IN
 CC DNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + (DNA)(N).
 CC -1- SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 80 kDa, 34 kDa, 30
 CC kDa, AND 29 kDa).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
 CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
 CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
 CC -1- SIMILARITY: TO OTHER SPECIES DNA POLYMERASE EPSILON, SUBUNIT B.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M61710; AAA34576.1; ALT_INT.
 DR EMBL: U25842; AAB68109.1; -.
 DR PIR: A39698; A39698.
 DR SDD: S0006379; DPB2.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KM DNA-binding; Nuclear protein.
 FT CONFLICT 461 461 F -> Y (IN REF. 1).
 FT CONFLICT 524 524 K -> R (IN REF. 1).
 FT CONFLICT 568 568 V -> P (IN REF. 1).
 FT CONFLICT 587 587 E -> Q (IN REF. 1).
 FT CONFLICT 647 647 T -> I (IN REF. 1).
 SQ SEQUENCE 692 AA; 78703 MW; 5C01647BD2B6A39A CRC64;

Query Match 2.5%; Score 109.5; DB 1; Length 692;
 Best Local Similarity 17.8%; Pred. No. 2.1; Mismatches 232; Indels 231; Gaps 31;
 Matches 123; Conservative 106;

183 VSRIKQEGVKMALHLFWFHPRVNSG-----FSIANRSTIKRLCTI----- 222
 Db 91 IQEKREKREKVMSEHRIQHEENILGRTDDENSDDEMPAADSSLOANVSLSSPMRPT 150
 QY 223 -----TESVEGEGSATL-----MSANPEGEPEDELMWQFVPAHLRASVSFLNFN 270
 Db 151 ERDEYKOPFKPSSKALDMRDYFKVINAQQOFRSPNPKHMQFIVENKKONGIGIAGF 210
 QY 271 LSNCKEER--VEYIIPGSTNPEVFKLEDKQGNAGNPN--LSLOGCDDAQSPGILR 327

```
Db 211 LPDIEDKQVOMFLTRYLL---TNDRYMRNENFNSDM---FNPLS-----SWS 252
Oy 328 LOFOVLVQHHPQMSNKIYVVDLSN-----ERAMSLTI 359
Db 253 LONELSNNTNRQOQSSMSSTPIKNLGRDAQNFLLLGLLNKNGWSLEDSGSVEIDI 312
Oy 360 EBRPVQSKHKEVPGCFVCLSERTCS-----SNLTITSGSKHKSIF-LCDLITRLMMNV 411
Db 313 SGTITPQGHYIYVPGCMVNLVEGIYISGNKEFHTYSMTLPRGKERREITLETIGNDLGITH- 371
Oy 412 EKTICTDHRVYQCRKRSYSLOVPDILH-LPVELHDFSMKLVPRK--DRLSLVLPAPKRL 467
Db 372 ----GISNNNFARLDKDLKIR---LHLEKEITDHKEFYILCANLFLDOLK-IMPLAKSI 423
Oy 468 QOHTHEKPCNTSFSLVASAIPSDLYRSGSPGSGSIKIQVKONISVTLRFPAPFOQE 527
Db 424 LQKLNDP-----PTLLIMQSF-----TSVPYFASM 450
Oy 528 ASROGLTVAFIYFKKE-----GVFTVTPDTKSKYVLRTPN--WDRGLPILTSVSN 577
Db 451 SSR-----NISSTOFKNNPDALATLSRFNLTENTMTITIPGNDLMG-----SMVSLGAS 503
Oy 578 ISVPDQVACLTFFKERSGVVCQGRAPMIIOEORTR---AEELFSLDEYVLPKPSFHH 633
Db 504 GLPDQDPIP--SAFTKKIKKVCKN---VWSSNPTRIAYLSOEIYIFRDLSGRKRRH 557
Oy 634 HSFWMNISCSPTSGKOLDLFSVTLTPRTVDTYLLAAGGVLLSALGLIICVCK 693
Db 558 LEFPNESDVTYENDNM-----575
Oy 694 KKKTKNKGPAGVIGYNGININTEMPROPKFQKGRKNDSHVAVIEDTMYGHLLODSGS 753
Db 576 MKKDIDIVI-----DELKPEPDQLEPOKVOETRK-----LVKTIIDOGH----- 614
Oy 754 FLQPEVDIYR---FQGTMGVCP-PSPTIC 780
Db 615 -LSPFLDSLRLPISWDLDTLTLCPPTMTWLC 645

RESULT 15
MAP2_MOUSE STANDARD: PRT: 1828 AA.
ID MAP2_MOUSE
AC P20357:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 2 (MAP 2).
GN MAP2 OR MAP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89083571; PubMed=3205744;
RT Wang D., Lewis S.A., Cowan N.J.;
RT "Complete sequence of a cDNA encoding mouse MAP2.";
RL Nucleic Acids Res. 16:11369-11370(1988).
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89043973; PubMed=3142041;
RT Lewis S.A., Wang D., Cowan N.J.;
RT "Microtubule-associated protein MAP2 shares a microtubule binding
RT motif with tau protein.";
RL Science 242:936-939(1988).
CC -1- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M21041; AAA39490.1; -.
DR PIR, S06467; S06467.
DR PIR, A40115; A40115.
DR MGD, MGI:97175; Map2.
DR InterPro, IPR001084; Tubulin_Tau.
DR Pfam, PF00418; tubulin-binding; 3.
DR PROSITE, PS00229; TAU_MAP_2.
KW Microtubules; Repeat; Calmodulin-binding.
FT DOMAIN 1452 1472 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 1662 1692 TAU/MAP MOTIF.
FT REPEAT 1693 1723 TAU/MAP MOTIF.
FT REPEAT 1724 1755 TAU/MAP MOTIF.
SQ SEQUENCE 1828 AA; 198980 MW; 200BC59E3360538CA CRC64;

Query Match 2.5%; Score 108.5; DB 1; Length 1828;
Best Local Similarity 18.1%; Pred. No. 11;
Matches 139; Conservative 108; Mismatches 285; Indels 237; Gaps 32;

Oy 228 EEGSATLMSANYPPEGPEDELTWQVYPAHLRASVFLN-----FNLSNCE-RKEER 280
Db 546 EAGSATIAEVEP--FYEDKSGMSKYFETSALEKEDMTSTELGSDIYELSDSRGSAQS 603
Oy 281 VEYIIPGSTNPVEFKLEDKQPGNMGAFNLSIQGDDQDAQSPGILRLQFOVLVQHHPNE 340
Db 604 LDTISPKNHDEKELQAKASQSPPAQEGYSTLA---QSYFPG---HPSELPEPSSP 656
Oy 341 SKTIYV-----DLSNERAMSLTIEPRPVKQSKRVYPCGVLCSERTCSNLTLS 391
Db 657 QERMFTIDPKVYGEKRDLSKKNKDLTL-----SRSGLGRSALIEGRMSINLPMSC 709
Oy 392 -----GSKHKISFLCDDTLRLMNVKTICTDHRVYQCRKYSLOVPDILH--- 438
Db 710 LDISALGFNFGCHDLSPLASDI-----LNTSGSMDGDDIYLPPTT 751
Oy 439 -----LPVELHDFSMKLVPRKRLSLVLPAPKLOQHTHEKPCNTSFSLVASAIPS 490
Db 752 PAVEKMPCFPIESKEEDEKAQAK-----VTCGGQIYQETSESPFAKEYKKNCTVMA 805
Oy 491 QDIYRSGSPGSGSIKQIQVKONISVTLRFPAPFOEASRQGLTVSFYFKKEGYFTYT 550
Db 806 PDLPEMLDLAGTRSRSLASVSADAEVARKRSVPS---EAMLAESSTL-LPVADESPYVK 861
Oy 551 PDKSKVYLRTPWMDRG-----LPSTLSVSNIS-----VPRDQVAC 587
Db 862 PDSQLE-----DMGYCVFNKTYVLPSPVODSENLSGSGSFYEGTDKVRDLAD 913
Oy 588 LTFFK-----ERSGV-----VCQGRAPM 606
Db 914 LSLIEVKLAAGRVKDEFAEKAETPTPSADKSGLSREFRDHKRANKDLDTYLEKSEHI 973
Oy 607 IIOEORTRAE-----ETISL-----DEVLPRPSFHHHSFWMNISCSPT 646
Db 974 DSEHAKESSEEMGKVELGLGITYDOASTKELITTKDTSPEKTEGLSSVEVAVEPT 1033
Oy 647 SGKQDLILFSVT-LPRTVDLTVYLLAAGGGVLLSALGLIICVKKKKKTKNKPAAG 705
Db 1034 TKADQGLDPAATKAEPSSQDLIKVSDRGMAAG-----AVDAGKALE 1075
Oy 706 I-----YNGNINTEMPROPKF-----QKGRKNDSHV-YAVIEDTMYGHLQ--- 748
Db 1076 LKFEVAQELTLESEAPQEADSEFWGESGHIKEGKVNETEVKRYKPLDLVQEAVDKBE 1135
Oy 749 --DSSSFLOPEVDITRPPQGTGVCPPSPPTICSAAPPAKATETPPRSPRESSEPY 806
Db 1136 SYESSGEHSLTMSLKPDGKRETSPT--SLIDQEVAKLSLVEIPCP--PPVSEAD-- 1189
Oy 807 TFSHPNNGV-----SKNDIDPL--NFOEPM-----EPAE 836
```

Tue Nov 12 18:03:31 2002

us-09-899-569a-4.rsp

Page 16

Db 1190 -LSTDEKGEVOMEFIQLPKKESTETPDIPAIPSDVTQPOPEAIVSEPAE 1237

Search completed: November 11, 2002, 11:31:35
Job time : 35 secs

.....